

**IJCSIS Vol. 13 No. 10, October 2015**  
**ISSN 1947-5500**

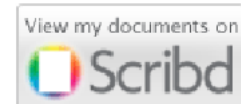
**International Journal of  
Computer Science  
& Information Security**

**© IJCSIS PUBLICATION 2015**  
**Pennsylvania, USA**



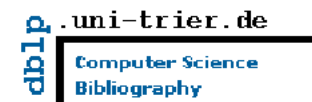
Cogprints

Google scholar



SciRate.com

CiteSeer<sup>x</sup> beta



DOAJ DIRECTORY OF OPEN ACCESS JOURNALS



ProQuest

# IJCSIS

ISSN (online): 1947-5500

Please consider to contribute to and/or forward to the appropriate groups the following opportunity to submit and publish original scientific results.

## CALL FOR PAPERS

### International Journal of Computer Science and Information Security (IJCSIS) January-December 2015 Issues

The topics suggested by this issue can be discussed in term of concepts, surveys, state of the art, research, standards, implementations, running experiments, applications, and industrial case studies. Authors are invited to submit complete unpublished papers, which are not under review in any other conference or journal in the following, but not limited to, topic areas.

See authors guide for manuscript preparation and submission guidelines.

Indexed by Google Scholar, DBLP, CiteSeerX, Directory for Open Access Journal (DOAJ), Bielefeld Academic Search Engine (BASE), SCIRUS, Scopus Database, Cornell University Library, ScientificCommons, ProQuest, EBSCO and more.

**Deadline:** see web site

**Notification:** see web site

**Revision:** see web site

**Publication:** see web site

Context-aware systems  
Networking technologies  
Security in network, systems, and applications  
Evolutionary computation  
Industrial systems  
Evolutionary computation  
Autonomic and autonomous systems  
Bio-technologies  
Knowledge data systems  
Mobile and distance education  
Intelligent techniques, logics and systems  
Knowledge processing  
Information technologies  
Internet and web technologies  
Digital information processing  
Cognitive science and knowledge

Agent-based systems  
Mobility and multimedia systems  
Systems performance  
Networking and telecommunications  
Software development and deployment  
Knowledge virtualization  
Systems and networks on the chip  
Knowledge for global defense  
Information Systems [IS]  
IPv6 Today - Technology and deployment  
Modeling  
Software Engineering  
Optimization  
Complexity  
Natural Language Processing  
Speech Synthesis  
Data Mining

For more topics, please see web site <https://sites.google.com/site/ijcsis/>

  
search engine for science

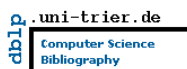




  
find and share professional documents

  
Bielefeld Academic Search Engine



  
Computer Science  
Bibliography

  
DIRECTORY OF  
OPEN ACCESS  
JOURNALS





For more information, please visit the journal website (<https://sites.google.com/site/ijcsis/>)

## Editorial Message from Managing Editor

*The International Journal of Computer Science and Information Security (IJCSIS) is a peer reviewed, high-impact online open access journal that publishes research which contribute new results and theoretical ideas in all areas of Computer Science & Information Security. The editorial board is pleased to present the October 2015 issue which consists of 22 high quality papers. The primary objective is to disseminate new knowledge and technology for the benefit of all, ranging from academic research and professional communities to industry professionals. It especially provides a platform for high-caliber researchers, practitioners and PhD students to publish completed research and latest development in these areas. We are glad to see variety of articles focusing on the major topics of innovation and computer science; IT security, Mobile computing, Cryptography, Software engineering, Wireless sensor networks etc. This scholarly resource endeavors to provide international audiences with the highest quality research and adopting it as a critical source of reference.*

*Over the last years, we have revised and expanded the journal scope to recruit papers from emerging areas of green & sustainable computing, cloud computing security, forensics, mobile computing and big data analytics. IJCSIS archives all publications in major academic/scientific databases and is indexed by the following International agencies and institutions: Google Scholar, CiteSeerX, Cornell's University Library, Ei Compendex, Scopus, DBLP, DOAJ, ProQuest, ArXiv, ResearchGate and EBSCO among others.*

*We thank and congratulate the wonderful team of editorial staff members, associate editors, and reviewers for their dedicated services to review and recommend high quality papers for publication. In particular, we would like to thank distinguished authors for submitting their papers to IJCSIS and researchers for continued support by citing papers published in IJCSIS. Without their continued and unselfish commitments, IJCSIS would not have achieved its current premier status.*

*"We support researchers to succeed by providing high visibility & impact value, prestige and excellence in research publication."*

*For further questions please do not hesitate to contact us at [ijcsiseditor@gmail.com](mailto:ijcsiseditor@gmail.com).*

*A complete list of journals can be found at:  
<http://sites.google.com/site/ijcsis/>*

*IJCSIS Vol. 13, No. 10, October 2015 Edition*

*ISSN 1947-5500 © IJCSIS, USA.*

*Journal Indexed by (among others):*







**Bibliographic Information**

ISSN: 1947-5500

Monthly publication (Regular Special Issues)  
Commenced Publication since May 2009

**Editorial / Paper Submissions:**

**IJCSIS Managing Editor**

[ijcsiseditor@gmail.com](mailto:ijcsiseditor@gmail.com)

**Pennsylvania, USA**

**Tel: +1 412 390 5159**

# IJCSIS EDITORIAL BOARD

Editorial Board Members	Guest Editors / Associate Editors
<b>Dr. Shimon K. Modi</b> <a href="#">[Profile]</a> Director of Research BSPA Labs, Purdue University, USA	<b>Dr Riktesh Srivastava</b> <a href="#">[Profile]</a> Associate Professor, Information Systems, Skyline University College, Sharjah, PO 1797, UAE
<b>Professor Ying Yang, PhD.</b> <a href="#">[Profile]</a> Computer Science Department, Yale University, USA	<b>Dr. Jianguo Ding</b> <a href="#">[Profile]</a> Norwegian University of Science and Technology (NTNU), Norway
<b>Professor Hamid Reza Naji, PhD.</b> <a href="#">[Profile]</a> Department of Computer Enigneering, Shahid Beheshti University, Tehran, Iran	<b>Dr. Naseer Alquraishi</b> <a href="#">[Profile]</a> University of Wasit, Iraq
<b>Professor Yong Li, PhD.</b> <a href="#">[Profile]</a> School of Electronic and Information Engineering, Beijing Jiaotong University, P. R. China	<b>Dr. Kai Cong</b> <a href="#">[Profile]</a> Intel Corporation, & Computer Science Department, Portland State University, USA
<b>Professor Mokhtar Beldjehem, PhD.</b> <a href="#">[Profile]</a> Sainte-Anne University, Halifax, NS, Canada	<b>Dr. Omar A. Alzubi</b> <a href="#">[Profile]</a> Prince Abdullah Bin Ghazi Faculty of Information Technology Al-Balqa Applied University (BAU), Jordan
<b>Professor Yousef Farhaoui, PhD.</b> Department of Computer Science, Moulay Ismail University, Morocco	<b>Dr. Jorge A. Ruiz-Vanoye</b> <a href="#">[Profile]</a> Universidad Autónoma del Estado de Morelos, Mexico
<b>Dr. Alex Pappachen James</b> <a href="#">[Profile]</a> Queensland Micro-nanotechnology center, Griffith University, Australia	<b>Prof. Ning Xu,</b> Wuhan University of Technology, China
<b>Professor Sanjay Jasola</b> <a href="#">[Profile]</a> Dean, School of Information and Communication Technology, Gautam Buddha University	<b>Dr . Bilal Alatas</b> <a href="#">[Profile]</a> Department of Software Engineering, Firat University, Turkey
<b>Dr. Siddhivinayak Kulkarni</b> <a href="#">[Profile]</a> University of Ballarat, Ballarat, Victoria, Australia	<b>Dr. Ioannis V. Koskosas,</b> University of Western Macedonia, Greece
<b>Dr. Reza Ebrahimi Atani</b> <a href="#">[Profile]</a> University of Guilan, Iran	<b>Dr Venu Kuthadi</b> <a href="#">[Profile]</a> University of Johannesburg, Johannesburg, RSA
<b>Dr. Umar Ruhi</b> <a href="#">[Profile]</a> University of Ottawa, Canada	<b>Dr. Zhihan Iv</b> <a href="#">[Profile]</a> Chinese Academy of Science, China
<b>Dr. Shimon K. Modi</b> <a href="#">[Profile]</a> Director of Research BSPA Labs, Purdue University, USA	<b>Dr Riktesh Srivastava</b> <a href="#">[Profile]</a> Associate Professor, Information Systems, Skyline University College, Sharjah, PO 1797, UAE

# TABLE OF CONTENTS

## **1. Paper 30091501: A Novel RFID-Based Design-Theoretical Framework for Combating Police Impersonation (pp. 1-9)**

*Isong Bassey & Ohaeri Ifeoma, Department of Computer Sciences, North-West University Mmabatho, South Africa*

*Elegbeleye Femi, Department of Computer Science & Info. Systems, University of Venda Thohoyandou, South Africa*

*Abstract* — Impersonation, a form of identity theft is recently gaining momentum globally and South African (SA) is not an exception. Particularly, police impersonation is due to lack of specific security features on police equipment which renders police officers (PO) vulnerable. Police impersonation is a serious crime against the state and could place the citizens in a state of insecurity and upsurge social anxiety. Moreover, it could tarnish the image of the police and reduce public confidence and trust. Thus, it is important that POs' integrity is protected. This paper therefore, aim to proffer solution to this global issue. The paper proposes a radio frequency identification (RFID) related approach to combat impersonation taking the South African Police Service (SAPS) as a focal point. The purpose is to assist POs to identify real POs or cars in a real-time mode. In order to achieve this, we propose the design of an RFID-based device having both tag and mini-reader integrated together on every PO and cars. The paper also implemented a novel system prototype interface called Police Identification System (PIS) to assist the police in the identification process. Given the benefits of RFID, we believed that if the idea is adopted and implemented by SAPS, it could help stop police impersonation and reduce crime rate.

*Keywords* — *impersonation, police, rfid, crime.*

## **2. Paper 30091502: An (M, K) Model Based Real-Time Scheduling Technique for Security Enhancement (pp. 10-18)**

*Y. Chandra Mouli & Smriti Agrawal, Department of Information Technology, Chaitanya Bharathi Institute of Technology, Hyderabad, India*

*Abstract* — Real Time Systems are systems where timely completion of a task is required to avoid catastrophic losses. The timely completion is guaranteed by a scheduler. The conventional schedulers only consider only the meeting the deadline as only design parameter and do not consider security requirement of an application. Thus modification of the conventional scheduler is required to ensure security to the Real time application. The existing security aware Real-Time scheduler (MAM) provides security to a task whenever possible and drops tasks whenever it is unable to schedule it within its deadline. The major problem in this tech scheduler is that it does not guarantee minimum security to all tasks. Thus, some may be exposed to security threats, which may be undesirable. Further, tasks are dropped in an unpredicted manner which is undesirable for Real Time Systems. This project presents an (M, K) model based Real Time scheduling technique SMK which guarantees that 'M' tasks in a window of 'K' successive task complete. It guarantees minimum security level to all the tasks and improves whenever possible. The simulation results show that the proposed SMK is approximately 15% better than the existing system.

*Keywords* — *Hard real-time scheduler, security, (M, K) model.*

### **3. Paper 30091503: Comparing PSO and GA Optimizers in MLP to Predict Mobile Traffic Jam Times (pp. 19-30)**

*W. Ojenge, School of Computing, TUK, Nairobi, Kenya*  
*W. Okelo-Odongo, School of Computing, UON, Nairobi, Kenya*  
*P. Ogao, School of Computing, TUK, Nairobi, Kenya*

*Abstract-* Freely-usable frequency spectrum is dwindling quickly in the face of increasingly greater demand. As mobile traffic overwhelm the frequency allocated to it, some frequency bands such as for terrestrial TV are insufficiently used. Yet the fixed spectrum allocation dictated by International Telecommunications Union disallows under-used frequency from being taken by those who need it more. This under-used frequency is, however, accessible for unlicensed exploitation using the Cognitive Radio. The cognitive radio would basically keep monitoring occupation of desirable frequencies by the licensed users and cause opportunistic utilization by unlicensed users when this opportunistic use cannot cause interference to the licensed users. In Kenyan situation, the most appropriate technique would be Overlay cognitive radio network. When the mobile traffic is modeled, it is easier to predict the exact jam times and plan ahead for emerging TV idle channels at the exact times. This paper attempts to explore the most optimal predictive algorithms using both literature review and experimental method. Literature on the following algorithms were reviewed; simple Multilayer perceptron, both simple and optimized versions of support vector machine, Naïve Bayes, decision trees and K-Nearest Neighbor. Although in only one occasion did the un-optimized multilayer perceptron out-perform the others, it still rallied well in the other occasions. There is, therefore, a high probability that optimizing the multilayer perceptron may enable it out-perform the other algorithms. Two effective optimization algorithms are used; genetic algorithm and particle swarm optimization. This paper describes the attempt to determine the performance of genetic-algorithm-optimized multilayer perceptron and particle-swarm-optimization-optimized multilayer perceptron in predicting mobile telephony jam times in a perennially-traffic jammed mobile cell. Our results indicate that particle-swarm-optimization optimized multilayer perceptron is probably a better performer than most other algorithms.

*Keywords – MLP; PSO; GA; Mobile traffic*

### **4. Paper 30091505: New Variant of Public Key Based on Diffie-Hellman with Magic Cube of Six-Dimensions (pp. 31-47)**

*Omar A. Dawood, Dr. Abdul Monem S. Rahma, Dr. Abdul Mohsen J. Abdul Hossen*  
*Computer Science Department, University of Technology, Baghdad, Iraq*

*Abstract -* In the present paper we are developed a new variant of asymmetric cipher (Public Key) algorithm that based on the Diffie-Hellman key exchange protocol and the mathematical foundations of the magic square and magic cube as the alternative to the traditional discrete logarithm and the integer factorization mathematical problems. The proposed model uses the Diffie-Hellman algorithm just to determine the dimension of magic cube's construction and through which determines the type of based magic square in the construction process if it is (odd, singly-even or doubly-even) type, as well as through which determined the starting number and the difference value in addition to the face or dimension number that will generate the ciphering key to both exchanged parties. From the other point it exploits the magic cube characteristics in encryption/decryption and signing/verifying operations. The magic cube is based on the folding six of series magic squares with sequential or with period numbers of n-dimensions that represent the faces or dimensions of magic cube. The proposed method speed up the ciphering and deciphering process as well as increases the computational complexity and gives a good insight to the designing process. The magic sum and magic constant of the magic cube play a vital role in encryption and decryption operations that imposes to keep as a secret key.

*Keywords: Magic Cube, Magic Square, Diffie-Hellman, RSA, Digital Signature.*

**5. Paper 30091507: Defining Project Based Learning steps and evaluation method for software engineering students (pp. 48-55)**

*Mohammad Sepahkar, Department of Computer Engineering, Islamic Azad University of Najafabad, Iran  
Faramarz Hendessi, Department of Computer & Electronic, Isfahan University of Technology, Iran  
Akbar Nabiollahi, Department of Computer Engineering, Islamic Azad University of Najafabad, Iran*

*Abstract* - Needing well educated and skillful workforce is one of the top items in industrial top priority list. But traditional education systems only focus on teaching theoretical knowledge to students which leads to lack of practical experience in them. Therefore modern pedagogy came to overcome this problem. Project based learning is one of these interactive learning pedagogies which is mostly used in engineering educations all over the world. In this research, we review a case study of executing a project based learning program in Isfahan University of Technology, Computer Engineering Department. During this overview, we explain all the steps needed for holding a PjBL curriculum with subject of software development. Finally we discuss about evaluation method for Project based learning programs.

*Keywords: Project based Learning, Education Pedagogy, Traditional Pedagogy, Software development, Team setup, Evaluation*

**6. Paper 30091511: Automated Recommendation of Information to the Media by the Implementation of Web Searching Technique (pp. 56-60)**

*Dr. Ashit kumar Dutta, Associate Professor, Shaqra University*

*Abstract* - Internet become the important media among the people all over the world. All other media depend on internet to gather information about the user navigational pattern and uses those information for their development. Web mining is the technology used for research carried out in internet. The notion of the research is to recommend the media to publish the frequently searched topics as news. The research uses google trends and hot trends data to find out the frequently searched topics by the user. An automated system is implemented to crawl items from the google trends and to recommend the same to the media.

*Keyword: Internet, Recommendation system, feeds, web mining, Text mining*

**7. Paper 30091513: Comparison of Euclidean Distance Function and Manhattan Distance Function Using K-Medioids (pp. 61-71)**

*Md. Mohibullah, Md. Zakir Hossain, Mahmudul Hasan, Department of Computer Science and Engineering, Comilla University, Comilla, Bangladesh*

*Abstract* -- Clustering is one kind of unsupervised learning methods. K-medioids is one of the partitioning clustering algorithms and it is also a distance based clustering. Distance measure is an important component of a clustering algorithm to measure the distances between data points. In this thesis paper, a comparison between Euclidean distance function and Manhattan distance function by using K-medioids has been made. To make this comparison, an instance of seven objects of a data set has been taken. Finally, we will show the simulation results in the result section of this paper.

*Keywords-- Clustering, K-medioids, Manhattan distance function, Euclidean distance function.*

## **8. Paper 30091520: Proposed GPU Based Architecture for Latent Fingerprint Matching (pp. 72-78)**

*Yenumula B Reddy, Dept. of Computer Science, GSU*

*Abstract* — Most of the fingerprint matching systems use the minutiae-based algorithms with a matching of ridge patterns. These fingerprint matching systems consider ridge activity in the vicinity of minutiae points, which has poorly recorded/captured exemplary prints (information). The MapReduce technique is enough to identify a required fingerprint from the database. In the MapReduce process, minutiae of the latent fingerprint data used as keys to the reference fingerprint database. The latent prints are analyzed using Bezier ridge descriptors to enhance the matching of partial latent against reference fingerprints. We implemented the MapReduce process to select a required document from a stream of documents using MapReduce package. MapReduce model uses parallel processing to generate results. However, it does not have the capability of using Graphics processing units (GPU) to execute faster than CPU-based system. In this research, we proposed a Python based Anaconda Accelerate system that uses GPU architecture to respond faster than MapReduce.

*Keywords:* fingerprint, minutiae points, MapReduce, Bezier ridge, GPU-based architecture;

## **9. Paper 30091523: Accelerated FCM Algorithm based on GPUs for Landcover Classification on Landsat-7 Imagery (pp. 79-84)**

*Dinh-Sinh Mai, Le Quy Don Technical University, Hanoi, Vietnam*

*Abstract* - Satellite imagery consists of images of Earth or other planets collected by satellites. Satellite images have many applications in meteorology, agriculture, biodiversity conservation, forestry, geology, cartography, regional planning, education, intelligence and warfare. However, satellite image data is of large size, so satellite image processing methods are often used with other methods to improve computing performance on the satellite image. This paper proposes the use of GPUs to improve calculation speed on the satellite image. Test results on the Landsat-7 image shows the method that authors proposed could improve computing speed faster than the case of using only CPUs. This method can be applied to many different types of satellite images, such as Ikonos image, Spot image, Envisat Asar image, etc.

*Index Terms-* Graphics processing units, fuzzy c-mean, land cover classification, satellite image.

## **10. Paper 30091525: Object Oriented Software Metrics for Maintainability (pp. 85-92)**

*N. V. Syma Kumar Dasari, Dept. of Computer Science, Krishna University, Machilipatnam, A.P., India.  
Dr. Satya Prasad Raavi, Dept. of CSE, Acharya Nagarjuna University, Guntur. A.P., India.*

*Abstract* - Measurement of the maintainability and its factors is a typical task in finding the software quality on development phase of the system. Maintainability factors are understandability, modifiability, and analyzability...etc. The factors Understandability and Modifiability are the two important attributes of the system maintainability. So, metric selections for the both factors give the good results in system of maintainability rather than the existed models. In the existing metrics obtained for Understandability and Modifiability factors based on only generalization (inheritance) of the structural properties of the system design. In this paper we proposed SatyaPrasad-Kumar (SK) metrics for those two factors with help of more structural properties of the system. Our proposed metrics were validated against the Weyker's properties also and got the results in good manner. When compare our proposed metrics are better than the other well-known OO (Object-Oriented) design metrics in getting the Weyker's properties validation.

*Keywords* – Understandability; Modifiability; Structural metrics; System Maintainability,; Weyker's properties; SK metrics; OO design;



**11. Paper 30091529: A hybrid classification algorithm and its application on four real-world data sets (pp. 93-97)**

*Lamiaa M. El Bakrawy, Faculty of Science, Al-Azhar University, Cairo, Egypt  
Abeer S. Desuky, Faculty of Science, Al-Azhar University, Cairo, Egypt*

*Abstract* — The aim of this paper is to propose a hybrid classification algorithm based on particle swarm optimization (PSO) to enhance the generalization performance of the Adaptive Boosting (AdaBoost) algorithm. AdaBoost enhances any given machine learning algorithm performance by producing some weak classifiers which requires more time and memory and may not give the best classification accuracy. For this purpose, We proposed PSO as a post optimization procedure for the resulted weak classifiers and removes the redundant classifiers. The experiments were conducted on the basis of four real-world data sets: Ionosphere data set, Thoracic Surgery data set, Blood Transfusion Service Center data set (btsc) and Statlog (Australian Credit Approval) data set from the machine-learning repository of University of California. The experimental results show that a given boosted classifier with our post optimization based on particle swarm optimization improves the classification accuracy for all used data. Also, the experiments show that the proposed algorithm outperforms other techniques with best generalization.

**12. Paper 30091532: Towards an Intelligent Decision Support System Based on the Multicriteria K-means Algorithm (pp. 98-102)**

*Dadda Afaf, Department of Industrial and Production Engineering, ENSAM University My ISMAIL, Meknes, Morocco  
Brahim Ouhbi, Department of Industrial and Production Engineering, ENSAM University My ISMAIL, Meknes, Morocco*

*Abstract* — the actual management of Renewable Energy (RE) project is involving a large number of stakeholders in an uncertainty and dynamic environment. It is also a multi-dimensional process, since it has to consider technological, financial, environmental, and social factors. Multicriteria Decision Analysis appears to be the most appropriate approach to understand the different perspectives and to support the evaluation of RE project. This paper aims to present an intelligent decision support system (IDSS), applied to renewable energy field. The proposed IDSS is based on combination of the binary preference relations and the multi-criteria k-means algorithm. An experimental study on a real case is conducted. This illustrative example demonstrates the effectiveness and feasibility of the proposed IDSS.

*Keywords-* Artificial Intelligence; Decision Support system, Multicriteria relation Clustering; k-means algorithm.

**13. Paper 30091533: Implementation Near Field Communication (NFC) In Checkpoint Application On Circuit Rally Base On Android Mobile (pp. 103-109)**

*Gregorius Hendita Artha K, Faculty of Engineering, Department of Informatics, University of Pancasila*

*Abstract* - Along with the rapid development of information technology and systems that were built to support business processes, then the required transaction data more quickly and safely. Several mechanisms are now widely used transactions with NFC include Internet Online Payment, Smart Cards, Radio Frequency Identification ( RFID ), Mobile Payment , and others. Where the mechanism - the mechanism is designed to simplify the user make transactions whenever and wherever the user is located. Build a new innovation from Checkpoint Apps In Rally Car circuits with Method NFC (Near Field Communication) Android Based Mobile. Basically , this is all the user system rally car competition organizers who set up several posts in the circuit for participants to be able to monitor the checkpoint that has been passed the participants are provided in each post - checkpoint . With the demand for speed in transactions , security , and ease of getting information , so the research is to discuss the checkpoint information on the rally car circuit method NFC ( Near Field Communication ) based mobile android . By using NFC technology in mobile devices connected to the checkpoint transaction process will be done faster, saving, and efficient. Application Circuit Rally Checkpoint On the Method of NFC (Near Field Communication) Android Based Mobile



can monitor the riders who are competing at a distance, so the crew team from each participating teams and the competition committee can see and track the whereabouts of the car which had reached a certain checkpoint. This application can be run through the android mobile to tell him where the car. The workings of web monitoring graphs are also features that can learn from each checkpoint and rally car so that it can be used easily in view of a moving car on the racing circuit. Android apps only support the devices that already have NFC reader , as in the designation as a liaison with NFC card . All mobile applications and websites related to the wifi network that has been provided so that the system can store data and display it on a website monitoring.

*Keywords- NFC, Near Field Communication, Android, Rally, Checkpoint*

#### **14. Paper 30091536: E-Government In The Arab World: Challenges And Successful Strategies (pp. 110-115)**

*Omar Saeed Al Mushayt, College of Business & Administration, KKU, Abha, KSA*

*Abstract* - Information Technology (IT) with its wide applications in all aspects of our life is the main feature of our era. It is considered as the telltale of development and progress of a country. That is why most countries are implementing IT in all areas through the establishment of the concept “e- government”. In this paper, we propose the importance of e-government, its contents, requirements, and then demonstrate the reality of e- government in the Arab World, discussing its challenges and successful strategies.

*Keywords: Information Technology, e-government, e-government in the Arab World.*

#### **15. Paper 30091530: CODMRP: A Density-Based Hybrid Cluster Routing Protocol in MANETs (pp. 116-122)**

*Yadvinder Singh, Department of Computer Science & Engineering, Sri Sai College of Engineering & Technology, Amritsar, India*

*Kulwinder Singh, Assistant Professor, Department of Computer Science & Engineering, Sri Sai College of Engineering & Technology, Amritsar, India*

*Abstract* — Cluster based on demand multicasting provides an efficient way to maintain hierarchical addresses in MANETs. To overcome the issue of looping in the ad hoc network, several ap-proaches were developed to make efficient routing. The challenge encountered by multicast routing protocols in this ambience is to envisage creating a cluster based routing within the constraints of finding the shortest path from the source and to convert a mesh based protocol into Hybrid. This paper represents a novel mul-ticast routing protocol C-ODMRP (Cluster based on demand routing protocol), a density-based hybrid, which is a combination of tree-based and mesh-based multicasting scheme. K-means algorithm approach also used to choose the *Cluster\_Head*, which helps in dynamically build routes and reduces the overhead of looping. C-ODMRP is well suited for ad hoc networks, as it choose *Cluster\_Head* through shortest path and topology changes frequently.

*Keywords-C-ODMRP, Cluster\_Head, K-means, density-based Hybrid, Route Table , MANETs.*

#### **16. Paper 30091526: Life time Enhancement through traffic optimization in WSN using PSO (pp. 123-129)**

*Dhanpratap Singh, CSE, MANIT, Bhopal, India*

*Dr. Jyoti Singhai, ECE, MANIT, Bhopal, India*

*Abstract* - Technologies used for wireless sensor network are extremely concentrated over improvement in lifetime and coverage of sensor network. Many obstacles like redundant data, selection of cluster heads, proper TDMA scheduling, sleep and Wake-up timing, nodes coordination and synchronization etc are required to investigate for the efficient use of sensor network. In this paper Lifetime improvement is an objective and reduction of redundant packets in the network is the solution which is accomplished by optimization technique. Evolutionary algorithms are one of the category of optimization techniques which improve the lifetime of the sensor network through optimizing traffic, selecting cluster heads, selecting schedules etc. In the proposed work the Particle Swarm optimization

Technique is used for the improvement in the lifetime of the sensor network by reducing number of sensor which transmits redundant information to the coordinator node. The optimization is based on various parameters such as Link quality, Residual energy and Traffic Load.

*Keywords: Lifetime, optimization, PSO, Fuzzy, RE, QL, SS*

### **17. Paper 30091521: Face Liveness Detection – A Comprehensive Survey Based on Dynamic and Static Techniques (pp. 130-141)**

*Aziz Alotaibi, University of Bridgeport, CT 06604, USA*

*Ausif Mahmood, University of Bridgeport, CT 06604, USA*

*Abstract* - With the wide acceptance of online systems, the desire for accurate biometric authentication based on face recognition has increased. One of the fundamental limitations of existing systems is their vulnerability to false verification via a picture or video of the person. Thus, face liveness detection before face authentication can be performed is of vital importance. Many new algorithms and techniques for liveness detection are being developed. This paper presents a comprehensive survey of the most recent approaches and their comparison to each other. Even though some systems use hardware-based liveness detection, we focus on the software-based approaches, in particular, the important algorithms that allow for an accurate liveness detection in real-time. This paper also serves as a tutorial on some of the important, recent algorithms in this field. Although a recent paper achieved an accuracy of over 98% on the liveness NUAA benchmark, we believe that this can be further improved through incorporation of deep learning.

*Index Terms* — *Face Recognition, Liveness Detection, Biometric Authentication System, Face Anti-Spoofing Attack.*

### **18. Paper 30091515: Cryptanalysis of Simplified-AES Encrypted Communication (pp. 142-150)**

*Vimalathithan. R, Dept. of Electronics and Communication Engg., Karpagam College of Engineering, Coimbatore, India*

*D. Rossi, Dept. of Electronics and Computer Science University of Southampton, Southampton, UK*

*M. Omana, C. Metra, Dept. of Electrical, Electronic and Information Engineering, University of Bologna, Bologna, Italy*

*M. L. Valarmathi, Dept. Computer Science, Government College of Technology, Coimbatore, India*

*Abstract* — Genetic algorithm based Cryptanalysis has gained considerable attention due to its fast convergence time. This paper proposes a Genetic Algorithm (GA) based cryptanalysis scheme for breaking the key employed in Simplified- AES. Our proposed GA allows us to break the key using a Known Plaintext attack requiring a lower number of Plaintext-Ciphertext pairs compared to existing solutions. Moreover, our approach allows us to break the S-AES key using also a Ciphertext-only attack. As far as we are concerned, it is the first time that GAs are used to perform this kind of attack on S-AES. Experimental results prove that our proposed fitness function along with GA have drastically reduced the search space by a factor of 10 in case of Known plain text and 1.8 in case of Ciphertext only attack.

*Index Terms* — *Cryptanalysis, Genetic Algorithm, Plaintext, Ciphertext, Simplified-AES.*

## **19. Paper 30091508: Risk Assessment in Hajj Event - Based on Information Leakage (pp. 151-155)**

*Asif Bhat, Department of Information Technology, International Islamic University Malaysia, Kuala Lumpur Malaysia*

*Haimi Ardiansyah, Department of Information Technology, International Islamic University Malaysia, Kuala Lumpur Malaysia*

*Said KH. Ally, Department of Information Technology, International Islamic University Malaysia, Kuala Lumpur Malaysia*

*Jamaluddin Ibrahim, Department of Information Technology, International Islamic University Malaysia, Kuala Lumpur Malaysia*

*Abstract* -- Annually, millions of Muslims embark on a religious pilgrimage called the “Hajj” to Mecca in Saudi Arabia. Management of Hajj activities is a very complex task for Saudi Arabian authorities and Hajj organizers due to the large number of pilgrims, short period of Hajj and the specific geographical area for the movement of pilgrims. The mass migration during the Hajj is unparalleled in scale, and pilgrims face numerous problems. Including RFID tags there are many types of identification and sensor devices developed for efficient use. Such technologies can be used together with the database systems and can be extremely useful in improving the Hajj management. The information provided by the pilgrims can be organised in the Hajj database and can be used to effectively identify individuals. The current system of data management is mostly manual, leading to various leaks. As more of the sensitive data gets exposed to a variety of health care providers, merchants, social sites, employers and so on, there is a higher chance of Risk. An adversary can “connect the dots” and piece together the information, leading to even more loss of privacy. Risk assessment is currently used as a key technique for managing Information Security. Every organization is implementing the risk management methods. Risk assessment is a part of this superset, Risk Management. While security risk assessment is an important step in the security risk management process, this paper will focus only on the Risk assessment.

*Keywords: Hajj, Information Leakage, Risk Assessment.*

## **20. Paper 30091527: Performance Evaluation of DWDM Technology: An Overview (pp. 156-172)**

*Shaista Rais, Dr. Sadiq Ali Khan*

*Department of Computer Science, University of Karachi*

*Abstract* - For abstract we shall discuss the following method. DWDM: dense wavelength division multiplexing. It is the method for expanding the data transfer capacity of optical system interchanges. DWDM controls wavelength of light to keep sign inside its own specific light band. In DWDM system dispersion and optical sign are the key elements. Raman and 100G advances are particularly discussed.

## **21. Paper 30091519: Information and Knowledge Engineering (pp. 173-180)**

*Okal Christopher Otieno*

*Department of Information Technology, Mount Kenya University, Nairobi, Kenya*

*Abstract* - Information and knowledge engineering is a significant field for various applications on processes around the globe. This investigation paper provides an overview of the status of development of the concept and how it relates to other areas such as information technology. The area that is of primary concern to this research is the connection with artificial intelligence. There is a revelation that knowledge engineering derives most of its operational domains from the principles of that concept. There is also a strong relation with the area of software development. As the research shows, they both have the same end products and procedures of attaining it. They both produce a computer program that deals with a particular issue in their contexts. The discussion also focuses on the two modeling approaches that are canonical probabilistic and decision based software processes. There is a description of the typical knowledge engineering process that each activity has to go through for efficient operation. The paper also takes a look at of the applications of knowledge-based systems in the industry.

## **22. Paper 30091518: Online Support Vector Machines Based on the Data Density (pp. 181-185)**

*Saeideh beygbabaei,*

*Department of computer Zanjan Branch, Islamic Azad University, Zanjan, Iran*

*Abstract* — nowadays we are faced with an infinite data sets, such as bank card transactions, which according to its specific approach, the traditional classification methods cannot be used for them. In this data, the classification model must be created with a limited number of data and then with the receiving every new data, first, it has been classified and ultimately according to the actual label (which obtained with a delay) improve classification model. This problem known the online classification data. One of the effective ways to solve this problem, the methods are based on support vector machines that can pointed to OISVM, ROSVM, LASVM. In this classification accuracy and speed and memory is very important; on the other hand, since finishing operations support vector machines only depends to support vector which is nearly to optimal hyperplane clastic; all other samples are irrelevant about this operation of the decision or optimal hyperplane, in which case it is possible classification accuracy be low. In this paper to achieve the desirable and accuracy and speed memory, we want by reflect the distribution density samples and linearly independent vectors, improve the support vector machines. Performance of the proposed method on the 10 dataset from UCI database and KEELS evaluation.

*Keywords:* *support vector machines, linear independent vector, relative density degree, online learning*

# A Novel RFID-Based Design-Theoretical Framework for Combating Police Impersonation

Isong Bassey and Ohaeri Ifeoma  
Department of Computer Sciences  
North-West University  
Mmabatho, South Africa

Elegbeleye Femi  
Department of Computer Science & Info. Systems  
University of Venda  
Thohoyandou, South Africa

**Abstract**—*Impersonation, a form of identity theft is recently gaining momentum globally and South African (SA) is not an exception. Particularly, police impersonation is due to lack of specific security features on police equipment which renders police officers (PO) vulnerable. Police impersonation is a serious crime against the state and could place the citizens in a state of insecurity and upsurge social anxiety. Moreover, it could tarnish the image of the police and reduce public confidence and trust. Thus, it is important that POs' integrity is protected. This paper therefore, aim to proffer solution to this global issue. The paper proposes a radio frequency identification (RFID) related approach to combat impersonation taking the South African Police Service (SAPS) as a focal point. The purpose is to assist POs to identify real POs or cars in a real-time mode. In order to achieve this, we propose the design of an RFID-based device having both tag and mini-reader integrated together on every PO and cars. The paper also implemented a novel system prototype interface called Police Identification System (PIS) to assist the police in the identification process. Given the benefits of RFID, we believed that if the idea is adopted and implemented by SAPS, it could help stop police impersonation and reduce crime rate*

**Keywords**—*impersonation, police, rfid, crime.*

## I. INTRODUCTION

Today, the world has witnessed a number of technological developments which has become widespread in all realms of life. Central to this is the exponential growth in the use of information and communication technologies (ICTs) that has proffered a platform for effective and efficient ideas, information and knowledge sharing [1][2]. In particular, the rapid proliferation of the Internet interconnectivity has changed the manner communication and businesses are performed whether personally, by organizations or the government [2][3]. While the derived benefits of the interconnectivity are great, they also poses significant risks that are known to be grievous and devastating [2]. In recent years, activities on the Internet has gained momentum as their physical life counterpart. Though, in the physical life a person is known to have one name to an activity, the case is not always the same on the Internet as one person can have several identities [4]. Consequently, the multiple identities can be used by such persons for the different purpose or services. This could be problematic. While there are several approaches or schemes in place to manage multiple identities online, multiple identities problems still exist in the physical life

today with negative impact on critical services delivery in the society. One of such issue is the continual impersonation of Police Officers (PO) and other uniform personnel in which the South Africa (SA) police is not an exception.

Impersonation is an illegal act which has gained global concern and nothing has been done to totally eliminate it. It is an act of stealing someone else's identity and assume the person's identity. Impersonation occurs when one person uses someone's personal information such as name, identity card, or credit card number, etc. to carry out actions not permitted such as frauds or other crimes. According to Marx [5], "...impersonation represents a kind of temporary identity theft that can hurt not only the duped, but society more broadly". This illegal act could be used to gain access to essential resources, services and other benefits in that person's name [6]. However, police impersonation is described by [7] as "...an act of falsely portraying oneself as a member of the police, for the purpose of deception". This deception carries great consequences as the impersonator tends to legitimize acts such as burglary, violent sexual assaults, robberies, killings, detaining [7][8], and so on. The offence class associated with police impersonation include verbal identification, fake badge, warrant card, fake uniform and fake vehicle [7][8]. These are used by the impostors to commit their crime under the police umbrella.

In several countries of the world, police impersonation is punishable with heavy custodial sentences attached. Several cases of police impersonation has been reported especially in the US, SA, Nigeria, Mexico and other countries of the world [8]. One of such cases is at the youth camp on the island in Norway on 22<sup>nd</sup> of July, 2011 where a man who posed as police officer started shooting at everyone [8]. While the impersonation act is easily accomplished, it is a serious crime that requires urgent and great attention. The fact remains that the police has an undisputed function of protecting lives and properties of the citizen of a nation. Hence, exploiting the vulnerability in the police may place the entire society at risk for easy harassment [8]. In addition, it could go a long way to quivering or reducing the confidence the public has in the law enforcement especially when they are oblivion of the real actors of a crime as impostors try to assert police-like authority. On this note, Marx [5] stressed that, "...unlike the

*current crime of identity theft, impersonating an 'agent of the state' is the theft or appropriation of a social identity*". Hence, the social implication is that, when confidence in the police get eroded, the citizens would be left in a state of insecurity and increased social anxiety [5][8]. This will then threaten the ability of the police to effectively perform their work, reduce level of trust and tarnish their reputation [5][8][9].

Today, police impersonation can easily be executed to commit a serious crime due to lack of distinct features associated with police equipment. Most people and even some members of the police authority have argued that those who impersonate the police are not harmful and their actions are being characterized as some form of tricks [10]. However, police impersonators are set to do more harm than just feigning to be POs. As the crime flourishes today, Callie et al [8] have suggested that the crime is common because of the insignificant penalties attached to the crime and nothing has been done to deter offenders. Therefore, our opinion in this paper is to rid impersonators in the society as their actions can promote insecurities and rid polices' trust and confidence. With the instruments that enables criminals to pose as PO, it is always difficult if not impossible to distinguish between real and nor real POs. Given the critical context, it is imperative that the police have to be protected from impostors in order to carry out their duties effectively. Though research on police impersonation is rare in the literature, this paper is geared towards offering a solution to the looming problem via the use of radio frequency identification (RFID) technology. The choice of RFID technology is that it has been known as one of the pervasive computing technologies that offers support for both practical and real-time implementation with reference to item identification, monitoring and tracking [18]. The central focus of this paper is the South African Police Service (SAPS) because of the high wave of police impersonation and other related crimes in the country. The integrity of the SAPS will be protected against impersonation vulnerability and consequently, if the RFID-based system is properly implemented, the confidence in the police will be strengthened.

The rest of the paper is organized as follows: Section II is the related works, Section III highlights impersonation crime in SA, Section IV is about RFID technology, Section V is the proposal, and Section VI is the proposed system operations. Accordingly, Section VII is the discussion, benefits and limitations while Section VIII is the paper conclusions.

## II. RELATED WORKS

RFID technology is gaining momentum in recent years and has received considerable attention more than other technologies in the automatic identification and data capture (AIDC) group. The technology has found application in various fields ranging from object, human or animal identification, tracing, tracking to monitoring [18]. Researches in the literature have shown that where RFID has been successfully applied, it yields positive impacts. Some of the successful applications of the technology are highlighted in this section:

One important application of RFID is in object monitoring. In this trend, an RFID-based livestock monitoring system was developed by [21]. The system monitors each livestock movement and also provides information about them using the RFID tag embedded on the animals. Moreover, in another study by [22], a Cold Chain monitoring system using RFID was developed and used to track the product movements in the supply chain. The system operates in a real-time mode where locations are tracked, temperature monitored to ensure products quality during delivery. In a similar work, a project called FARMA was developed by [23]. It uses the RFID technology alongside wireless mobile network to track animals as well as access their information stored in a data repository. The basis was to track and identify the animals when they gets lost.

In another successful RFID application by [24], a context aware notification system based on RFID was designed and developed for university students. The notification system was developed with the goal of delivering urgent notifications to expected students instantly where there were at the moment according to Haron et al [24]. Furthermore, Herdawatie et al [25] designed a student tracking system that track students' location in a boarding school using a combination of RFID and biometric sensor. In the same vein, Christopher et al [26] also developed a system that automate the long-term mice behavior watching in socially or structurally complex caged environments via RFID system. The system accurately accounts for the locations of all mice as well as each mouse's location information over time and so on. In akin work by [27], an RFID-based system was developed to identify patients in the hospital. The aim was to identify patients faster especially in the case of unconscious patients or those that can't communicate which could delay treatments. Also, Catarinucci et al [28] developed an RFID-based automated system that tracks and analyses behavior of rodents in an ultra-high-frequency bandwidth.

Highlighted above are some of the successful application and there exist several others we have not discussed in this paper. Based on these successes in tracking, monitoring and identifying, we strongly support that the application of RFID in combating police impersonation could assist SAPS in identifying who is a real PO and who is not.

## III. POLICE IMEPERSONATION IN SOUTH AFRICA

In our society today, identity theft is not a new phenomenon as it has remained in existence for some times now and impersonating the police is not an exception. However, when this becomes widespread among law enforcement agents, insecurity and increased social anxiety could become the order of the state [5][8]. Considering the primary functions of the police which include protecting the citizenry from crime, maintaining law and order, preventing and controlling terrorist activities and any other threats that can undermine the peace and harmony of a nation, there is need for the police to be protected from any form of threats that can place the society at



risks of insecurity and lack of trust and confidence on the part of the police.

With the focal point on SAPS, several incidences of police impersonation has increased in recent years and some have been reported on the media. Among such cases is one reported by [11], involving a convicted rapist and a thief who ran away from jail and later became a police captain in Polokwane police station. It was embarrassing that he was never checked as a valid PO and record never assessed but was allowed to performed full police functions. Another related incident is the case of police impostors in KwaZulu-Natal (KZN) where it was alleged criminals broke into POs home, stole their uniforms and cloned their identification cards [12]. According to the report, the equipment were then used for armed robberies, with well-known businessmen as their target. The extorted money from their victims but were later arrested by the police. Several other cases of police impersonation have also been reported in KZN where uniforms and identification cards were used. However, the police authority argued that people posing as POs was possible because it was easy to clone their identity card since "...there were no special features on it"[12].

Furthermore, on the 26th of February, 2015, another case involving four criminals that impersonated Johannesburg metro POs was reported [13]. The criminals were caught with bulletproof vests and fake appointment cards. The motive was solely to commit crime and fraud. Similarly, on the 28th of July, members of the police unit called the Hawks arrested seven criminals that impersonated the Hawks to extort money from their victims [14]. In another event, on the 9th of August, 2013, the police recover a police car that was reproduced or cloned and was believed to be involved in 35 separate cases of robbery, hijacking and theft [15]. (see Fig. 1) They were caught with handguns, rifle and a bulletproof.



Figure 1. Police recover cloned flying squad car [15]

The above highlighted crimes are some of the cases of police impersonation in SA. Although the actual estimates are not represented in this paper, the country have experienced stern incidents of police impersonation and if allowed to continue, it could lead to severe criminal behaviour. Furthermore, if it is not stopped, the public will no longer repose their confidence in the SAPS. Consequently, it will have negative implications on the state and the citizenry [5]. The fact is that, posing as a police allows criminals to operate freely without being challenged by law enforcement agents, leading to increased crime rate, reduced police trust and so on. The onus rest in the

hands of the SAPS to eliminate the menace. This therefore, constitute the motivation for this paper. Our aim is to propose a cost-effective way of fighting police impersonation in SA.

#### IV. RFID TECHNOLOGY

RFID is an electronic device that is used to uniquely identify an item or person automatically and wirelessly via radio wave [16][17]. It is consist of a small chip and an antenna which can automatically identify, track, and store information. At minimum, the RFID system components are the tags, the readers and the middleware application which integrated into a host system that processes the data [16]. The tag stores the information about an object and the readers are used to capture data on tags and send to the computer system remotely automatically. (See Fig. 2)

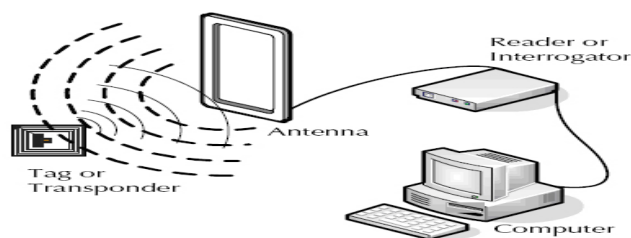


Figure 2. RFID System

RFID technology has been in existence for more than eighty years and is in the Automatic Identification and Data Capture (AIDC) technology category [18][29]. For effective communication of RFID-related activities, the EPCglobal Network is used. It is a suite of network services that is developed to share products or objects data which are RFID-related in the world through the Internet. The network was developed by the AIDC but currently managed by EPCglobal Inc [29][30]. The network seats on several technologies and standards in which Electronic Product Code (EPC) is the basis for information flow. EPC is considered to be a universal identifier in which a particular physical object is given a unique identity [30]. EPC is stored on the RFID-tag which is scanned by the reader. Other key components of the network includes: the Object Naming Service (ONS), the EPC middleware or savant, the EPC Information Service (EPCIS), and the EPC Discovery Services (EPCDS) [30]. (see Fig. 9) Also basic functions of each component are discussed. The ONS constitute a service that helps in information discovery of an object using the EPC. Upon a given information request or query, ONS obtain the EPC and yield a matched URL where the object information is stored in the database. The EPC savant has the task of collecting RFID tag data from the reader, filter and aggregates tag data and passes processed tag data to the application [30]. EPCIS is simply the EPC database which performs the services of the storage, hosting and enable the sharing of a particular product information that is EPC-related. Accordingly, the EPCDS perform the services of tracking and tracing efficiently. In this case, it keeps record of each EPCIS with instance of a particular object's data. For more details about RFID and how it operate, refer to [20][30].



RFID technology has proffer numerous benefits which are not possible with other members of the AIDC technologies. It includes reduced automatic and wireless identification, object tracking and tracing, data accuracy improvement, reduction in time and labour for manual data input and so on. Moreover, it has attracted a wide application areas such as inventory tracking, logistic and chain supply, racing timing, access control, asset tracking, real time location systems, patient's identification, and so on [18][19]. Based on the successful application of RFID both in the industry and academia, we strongly support that its incorporating into police equipment would go a long way to assist stem the tides of the risks posed by police impersonators in the SA society.

### V. THE POPROPOSAL

The police play a critical role in every nation of the world. However, they are not immune to attacks and impersonation. In particular, police are impersonated by criminals to commit serious crime. According to [12], this could be as a result of lack of strong security measures on police equipment to authenticate and validate that a person wearing such a uniform or in possession of other police equipment is a real PO. The consequence of this has resulted in high rate of crimes worldwide. Thus, it is important that POs are protected from impersonation crime to enable them perform their tasks effectively and to reduce crime rate in the society to the barest minimum.

Therefore, in order to protect the integrity of the police, this paper proposes the design of a security system that is based on RFID technology that utilizes EPCglobal Network to solve the impersonation among the PO in SA. To achieve the system design, requirements were gathered using the observation technique which involved taking an in-depth look at police and the impersonators' mode of operations. The overall requirements that need to be satisfied by the system is the identification of PO as real or not real. Details about the system architecture is discussed in the sub-sections that follows.

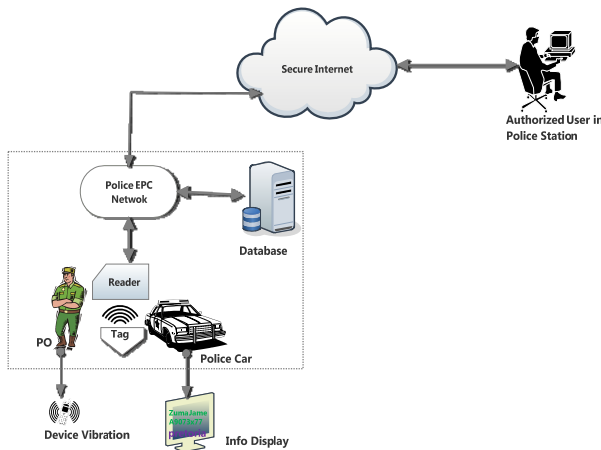


Figure 3. Proposed system architecture

### A. System Architecture

The architecture of the system consist of several components which are: (1) RFID tag and reader, the (2) Police central database, (3) the Police EPCglobal Network and the Police computer. The overall structure of the system in terms of identification and authentication as well as their interactions is shown in Fig. 3. With the system mode of operation, there is the requirement that the police maintain a central database where information about real POs as well the information about police cars are collected and stored. Fig. 4 shows the relational diagram of the database containing information to be collected for both cars and POs.

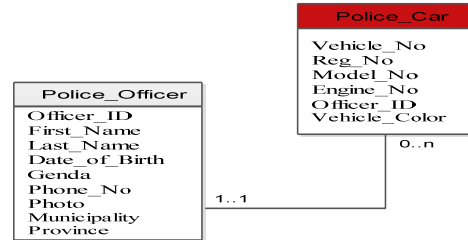


Figure 4. Relational diagram

To facilitate the identification process, each PO will be assigned a unique identifier called **Officer\_ID** and **Vehicle\_No** for each police car. In the same vein, each PO will be given an electronic device that is integrated with RFID tag and mini-reader. The tag will contain the EPC code which is the **Officer\_ID** or the **Vehicle\_No**. The RFID reader will read the tag data wirelessly which will be used to identify if a PO or car is real or not. This process will be made effective in a real-time manner by utilizing the EPCglobal Network [30]. If an officer is identified as real PO, a **DEVICE VIBRATION** will be observed, otherwise is an impersonation. For the police car, information will be displayed on the screen of the computer installed inside the police car, otherwise is a cloned police car. The illustration of the scenarios is captured in Fig. 10 and 11 respectively.

### B. RFID Tag and Reader

For effective communication and in line with POs' operation, the design of an electronic device that will contain both the tag and reader integrated together is recommended. That is, an active RFID tag and a mobile RFID reader. The mobile RFID reader will have the capability to scan and wirelessly sends tags information via the EPC network for processing and receives the results that identifies a PO to be real or not irrespective of the location. In addition, the structure of the RFID tag will be in line with the EPCglobal specification which is shown in Fig. 5.

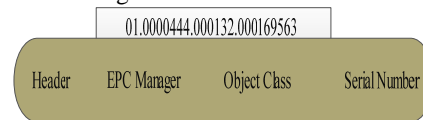


Figure 5. Police RFID tag structure

**Header:** In the device, the header will identify the type, length, structure, version that these tags will take, which could be 64bit, 96bit and 256 bit, as needed.

**EPC Manager:** The will contain the number that identifies an organizational entity. In this case, the manager will have a code that represents the SAPS who is responsible for the device the EPC is embedded to.

**Object Class:** It identifies the exact type of product. In this case, the object class will identify a specific province where each PO's information and the police car's information belong.

**Serial Number:** This is a unique number of items within each object class. The serial number will be the unique identification number of each PO. The number will be used to query the central databases through the secured Internet to retrieve, update, identify and authenticate individual stored POs/assets' information.

### C. RFID Reader and Central Database Interactions

Based on the design that incorporates both tag and scanner, care have to be taken to ensure that each RFID reader do not read its own tag data every time it senses a tag. To this end, each RFID reader will be embedded with the intelligence of identifying its own tag data. This will be achieved by having the reader and the tag must store the same serial number. Hence, the algorithm represented by the flowchart captured in Fig. 6 can be followed to design the device that contains tag and the reader. As indicated in the algorithm, anytime the reader scans a tag, it gets the tag data first and compares it with its own code. If they are the same, it does nothing, otherwise it uses the tag data to query the police database in order to identify and authenticate if a PO or police car is real or not.

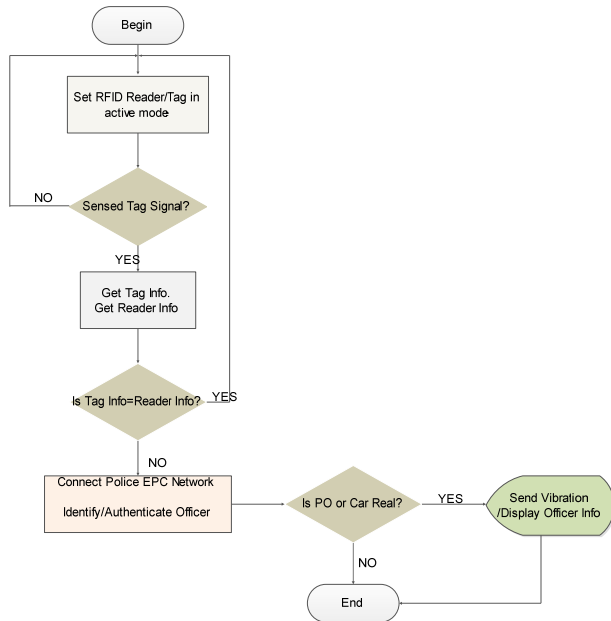


Figure 6. RFID reader and tag communication

In the same vein, the RFID reader must be able to communicate with the police central database on a real-time mode in order to carry out the task of identification and authentication effectively and efficiently. The communication is represented using the sequence diagram as shown in Fig.7.

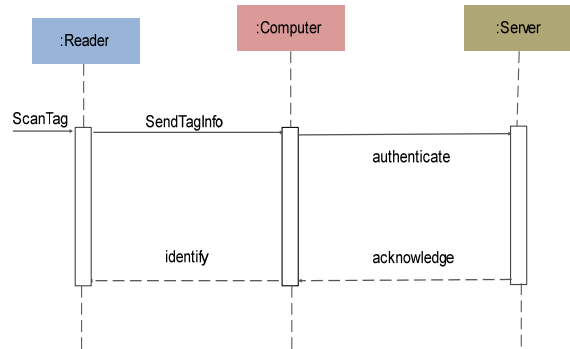


Figure 7. RFID scanner/Central database sequence diagram

As underlined in Fig. 7, in order to identify and authenticate a PO or police cars as real or not, the RFID reader will read the tag data (Officer\_ID or Vehicle\_No) and transmit tag data via the EPCglobal Network. The tag data is then automatically used as a primary key to query the central database. If the tag data matches information stored in the database, an acknowledgement is sent in the form of vibration or information display on the computer screen. Otherwise, it is a case of police impersonation or car clone. The architecture of the authentication and identification process is shown in Fig. 8.

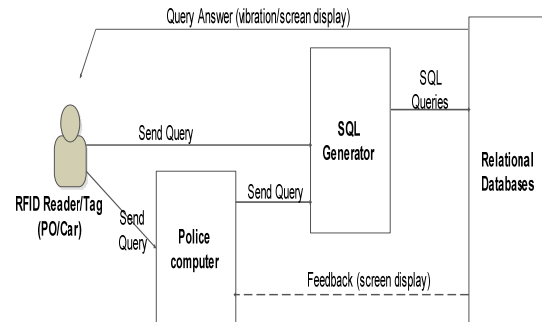


Figure 8. Architecture for authentication and identification query

### D. Police EPCglobal Network

EPCglobal Network provides several benefits which forms part of the choice of its applicability in this work. Some of the benefits are: (1) the network is designed to provide a link for all EPC tag-oriented physical objects, (2) scale to support huge volume of data generated by RFID-enabled activities between readers and tags, and (3) to proffer a data format that is universal for transferring information specific to a particular product or object [29][30]. This work thus, take advantage of the benefits offered by the technology and extend it to combating impersonation of POs and cloning of police cars. The EPCglobal Network architecture is captured in Fig. 9.

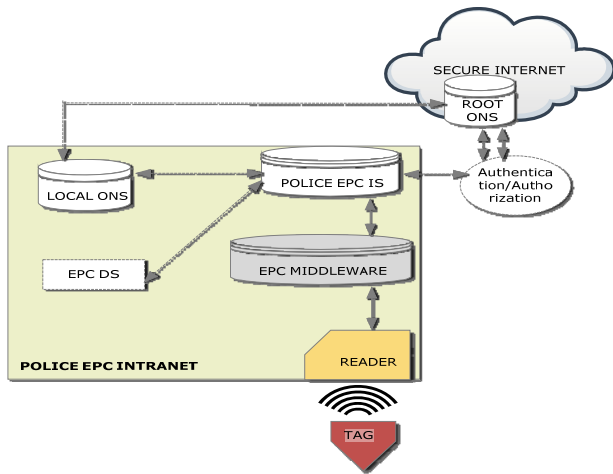


Figure 9. Police EPCglobal network architecture

Based on the function of each component as discussed in Section IV, the contribution of each to the effective operation of the proposed system is as follows within the Police EPC intranet. Given the electronic device having a tag and a mini-reader on a PO or a police car, in the event that an RFID reader reads an RFID data, the data which is the EPC will then be communicated by the reader through the EPCglobal Network to the middleware. The RFID savant which has the capability to process, filter, aggregate the tag data will in turn, use the processed EPC to query the root ONS over the secured Internet for the Police local ONS. At this point, the root ONS queries the local ONS for the location of the PO or Car data which is in the EPCIS. To access the data for a particular PO or car, the EPCIS is queried using the EPC-related data and if a matched is found in the database maintained by EPCIS, the result of the request is sent back to the device in the form of a vibration or screen display, confirming the PO or car is real. Lastly, the EPC DS will keep track or record of all police cars or PO identified by EPCIS.

## VI. SYSTEM OPERATION

In this section, we discuss the operation of the proposed system. However, for the system to operate effectively and to meet the overall goal of protecting the integrity of the police, different scenarios are explored where POs or police car could be vulnerable to impersonation. The different scenarios are as follows:

- Officer-to-Officer identification
- Officer-to-Car and Car-to-Car identification

These scenarios are intended to proffer solution to the identification of real POs or real police cars by the proposed system to thwart police impersonation. The justification is that today for instance, people can have access to police uniforms and clone police cars which they used to commit serious crimes. Moreover, it is always difficult to know who is real or what is real. Therefore, it is important to ensure that only real POs are allowed to perform the tasks of policing and protecting lives, properties and other essential functions. The operation of each scenario is discuss as follows:

### A. Officer-to-Officer Identification

In order to identify that an officer is a real PO, let assume two POs **A** and **B** who are unknown to each other get in physical contact either on the street or at a police checkpoint on the road. This system is poised to assist the POs to identify themselves secretly as real or not real. The identification process is demonstrated in Fig.10. In this case, each PO is given a wristwatch-like electronic devices that contain both the RFID tag and a mini-reader. As officer **A** approaches officer **B** and both are within the frequency range of the RFID reader, each reader will automatically and wirelessly read the tag data and transmit it via the EPCglobal Network to query the central database using the *Officer\_ID*. Upon a match, the officer will be authenticated and receive acknowledgement in the form of a *vibration*. This will confirm that he or she is a real PO, otherwise, is a police impersonator.

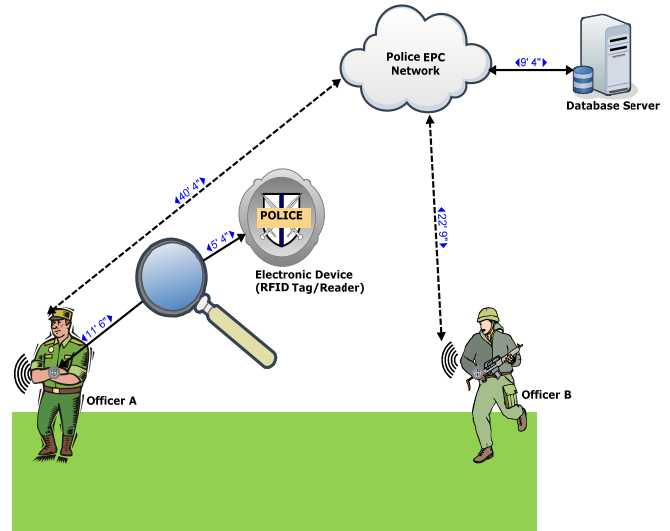


Figure 10. Two police officer identifying themselves

### B. Officer-to-Car and Car-to-Car Identification

In another scenario where police impersonators can be identified is a situation where impersonators use a cloned police car. In order to identify impersonators in this direction, the operation is captured in Fig. 11. Assuming that a police car say, **Car A** is in physical contact with another police car say, **Car B** or POs in the street or a checkpoint mounted on the road. They have to first identify each other at a reasonable distance in order to prepare in advance for any action in the event of impersonation. This is necessary to effectively and efficiently protect the officers on the checkpoint or in the car. With this system, for the POs on the checkpoint or in police **Car A**, as the police **Car B** approaches the checkpoint and is within the frequency range of the identification device, the mini-RFID reader will automatically read the tag data and transmit the data via the EPCglobal Network to query the central databases.

- For the officers at the checkpoint, if the car is a real police car, say **Car A**, the database will return an

acknowledgement in the form of a vibration on the device on the officer's wrist.

- For the patrolling **Car** say A or B, if the PO at the checkpoint is a real PO, the database will return an acknowledgement by displaying the PO's information on the screen of the computer installed inside the car.
- For police car **A** and **B**, if **Car A** approaches **Car B**, then the databases will return acknowledgement by displaying each registered police car's information on the screen of the computer. The information that will be displayed if it is found to be a real police car is shown in Fig. 4.

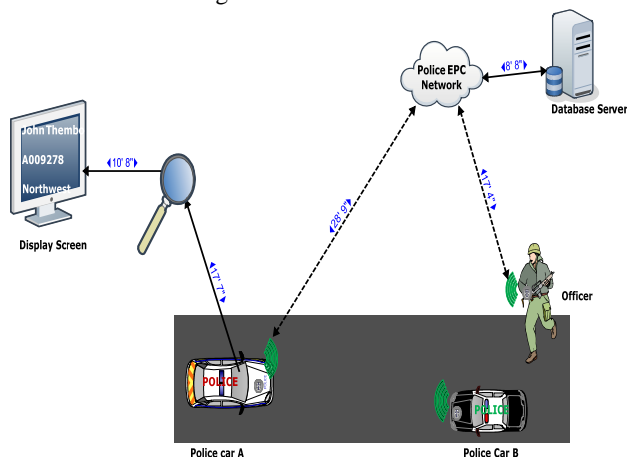


Figure 11. PO and car identification

However, in whichever way the identification takes place, the display of information or the receipt of vibration will prove that the officer or the van is a real PO or car. If no information is displayed or vibration felt, it proves that the police has been impersonated. Therefore, measures have to be taken to ensure that those involved in such act are brought to justice.

### C. Police System Interface

This section presents the system prototype called *Police Identification System* (PIS) that implements the proposed system interface. For effective usage and information processing, the software system will be installed on the police computer or mobile devices either in the police station or police cars to assist them in carrying out any task of registration, identification and validation of police officers. The system shall allow POs to login, register, search, print, display, and update POs' information. However, due to some constraints, this work is only at the proposal stage where only the design of the RFID tag, reader and the network have been discussed. Moreover, we have only implemented system prototype interfaces which are shown in this section.

1) *System administrator interface*: The system administrator is responsible for the registration and management of the PO information stored in the database. However, in order to get to the PIS home page, only

administrator with valid credentials will have access to the system. The home page has all the functionalities the administrator can perform such as registration, scanning, displaying of each officer and police's car details into and out of the databases. (See Fig. 12)

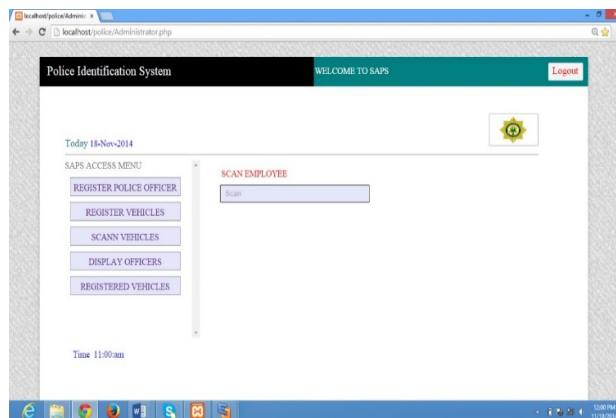


Figure 12 PIS home

2) *Police vehicle registration and identification interface*: This interface will assist the administrator to capture police cars' information into the police database. The data to be collected are specified in Fig. 4. The interface that enables the capability is show in Fig. 13.

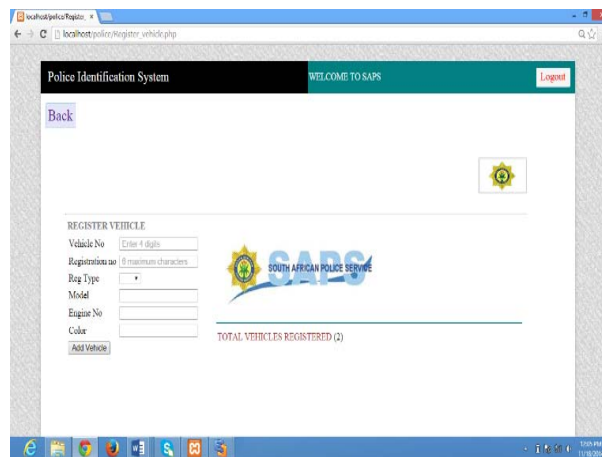


Figure 13. PIS Van registration

In addition, the stored information can be used to identify police cars in a real-time manner by the interaction between the tag, reader, EPC Network and the central database. (See Fig. 14). In the event of any interaction, if a tag data matches the data in the police database, the car's information will be displayed, indicating a real police car. In the same vein, an office can also search for a police car using the unique *Vehicle\_No*.



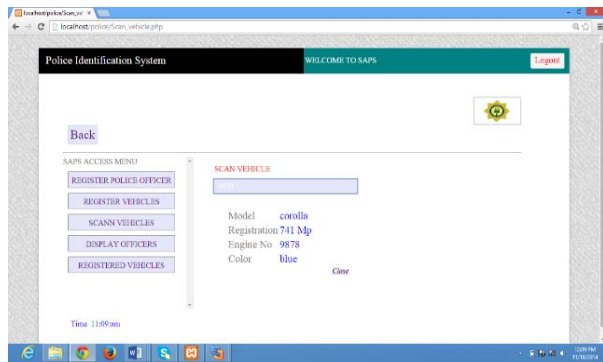


Figure 14. Police car identification

3) *Police registration and identification interface:* Just like the police car, all POs will be registered for onward identification, when the need arises. In this case, the essential information about every PO will be captured and stored in the central database. The registration form is shown in Fig.15 and has the capability of updating officers' information regularly. Also, Fig. 16 is the structure of table that contains already registered POs' information.

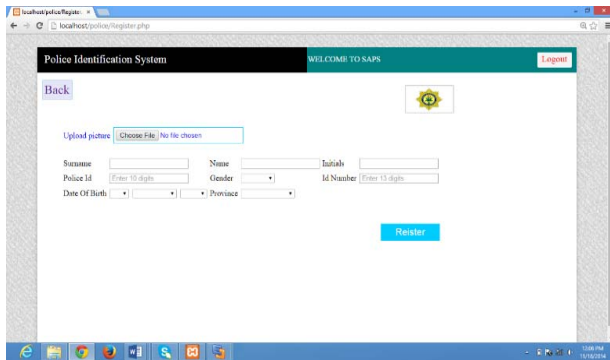


Figure 15. PIS registration form

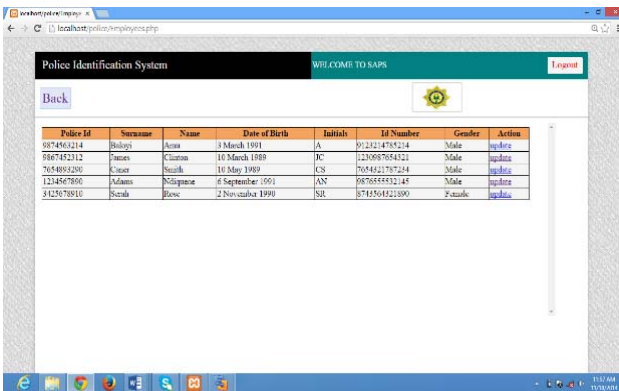


Figure 16. PIS Registered police officers table

However, the identification of POs has to be done on a real-time mode just like the police car. In this case, if the tag data

matches the data in the police database, the officer's information will be displayed automatically as shown in Fig. 17. This will prove that the PO is a real, otherwise, an imposter.

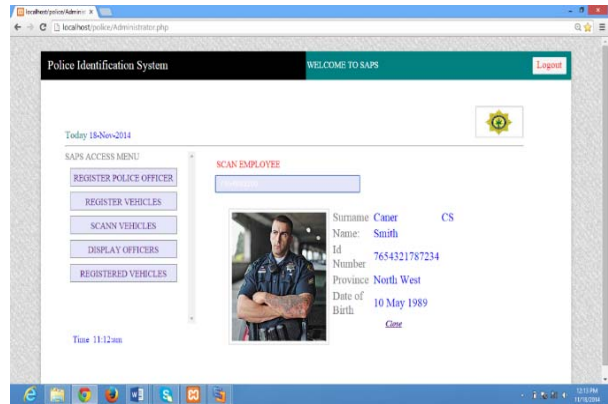


Figure 17. Identified real PO

In this section, we have presented few of the interfaces of the prototype showing some of the system functionalities in its prototypic phase. A complete system will be implemented when the idea discussed in this work is subscribed for by the appropriate authority.

## VII. DISCUSSION, BENEFITS AND LIMITATIONS

In this paper, we have proposed and discussed an approach that will help curb impersonation in the SAPS. The objective is to ensure that the police integrity is protected in a way that the citizens would always have confidence in police and carrying out their duties effectively. To achieve this, we have proposed an RFID-based system designed in the form of a wristwatch, having both tag and mini reader that can communicate with the database in a real-time manner via the EPCglobal Network. The system is to assist POs identify any person in police uniform or cars posing as real POs or cars. Additionally, the system operations has also been discussed in Section V and as a proof of concept, a prototype called PIS was implemented, having interfaces that would assist POs interact with the system effectively. With PIS, officers can register every police officer into the system for identifications. Therefore, based on the system operation, we believe that if accepted for use, it would go a long way to provide important benefit to the general police force in SA and the African continent at large to curb impersonation and crime rate. Some of the benefits are to reduce crime rate that emanates from police impersonation, increase security, eliminates social anxiety, strengthen polices' ability to do their job and increases confidence and trust in the police force. Lastly, the creation and maintenance of databases which will operates in a real-time basis could also be used for dual purposes that is beneficial to both government and non-governmental organizations.

However, in spite of the enormous benefits our system is without limitations. The limitations are: (1) if the tags is not

properly secure it could easily be stolen and used for advance impersonation to commit more serious crimes and frauds, (2) being a web-based system, it could also be subjected security attacks such as hacking, denial of service, etc. (3) Unavailability of electricity or network can affect the system adversely, (4) it could be difficult or impossible to realize the design of the tag and reader and their cost might hinder the system implementation, and (5) the system does not in any way provide for the general public to identify who is a real PO which could be a serious issue. These and other limitations not mentioned in this paper could affect the smooth functioning of this system negatively. However, necessary measures will be taken to address them in the event the system is adopted for usage.

### VIII. CONCLUSIONS

Security is critical to an individual, organizations and nations. As various forms of crimes are being experienced on a daily basis, there is need for strategies in place to reduce the menace. In recent years SA has witnessed an upsurge in the number of police impersonation leading to high crime rates and safety concerns in the country. Though some of the impostors were apprehended, there is need to get rid of this crime and protect the police in order to carry out their duties effectively. In this paper, we have proposed a technique which could serve as a solution to the existing challenges faced by SAPS. The technique make used of RFID technology to remotely identify and authenticate PO as real. In addition, we developed a system prototype showing various interfaces offered by the system. We also discussed the benefits that can be derived from adopting system as well as its limitations. Based on the system operation, we therefore conclude that if this system is accepted for used in the SAPS or other related security agencies, it could go a long way to provide citizens with more security and get rid of all forms of social anxiety. Accordingly, it will boost the public confidence in the police force and increased trust. This proposed system would serve as a foundation to solving impersonation problems in the police. The future work will be to test the operation of the system in the real-world object, develop a complete system and present it to SAPS for evaluation and onward usage.

### REFERENCES

- [1] Welch, E. W and Hinnant, C.C., "Internet Use, Transparency, and Interactivity Effects on Trust in Government" Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03) IEEE, 2003
- [2] Nordwood, K.T. and Catwell, S.P. Cybersecurity, Cyberanalysis, and Warning, Nova Science Publishers, Inc. 2009
- [3] Andreasson, K. Cybersecurity: Public Sector Threats and Responses CRC Press, Taylor & Francis Group, 2012.
- [4] Han, B. et al. BioID: Biometric-Based Identity Management. ISPEC 2010, LNCS 6047, pp. 241–250, 2010.
- [5] Marx, G. Are you for real? Police and other impersonators, 2005. Retrieved from <http://web.mit.edu/gtmarx/www/newsday11605/html>
- [6] Ali, H. "An analysis of identity theft: Motives, related frauds, techniques and prevention". DOI:10.5897/JLCR11.044.ISSN2006-9804. 2012
- [7] Wikipedia, Police impersonation. Retrieved from [http://en.wikipedia.org/w/index.php?title=Police\\_impersonation&oldid=651408514](http://en.wikipedia.org/w/index.php?title=Police_impersonation&oldid=651408514)

- [8] Rennison, C.M and Dodge, M. Police Impersonation: Pretenses and Predators. American Journal of Criminal Justice, Dec. 2012, Vol. 37 Issue 4, p505
- [9] Tyler, T. H. (2004). Enhancing police legitimacy. The ANNALS of the American Academy of Political and Social Science. 593, 84–99.
- [10] Van Natta, D. (2011). In Florida, criminals pose as police more frequently and for more violent ends. Retrieved from <http://www.nytimes.com/2011/05/29/us/29fakecops.html?>
- [11] S., David. Fake South African police officer on the run after almost three years on the beat, The Guardian. Monday 23 February 2015. <http://www.theguardian.com/world/2015/feb/23/south-african-fake-police-officer-limpopo-capt-mailula>
- [12] Barbeau, N and Pillay.K. Police impersonation on the rise South Africa Daily New Report, 2012. <http://www.iol.co.za/dailynews/news/police-impersonation-on-the-rise-1.1428630#>
- [13] SAPA, Four held for impersonating JMPD, iolnews. February 26 2015. <http://www.iol.co.za/news/crime-courts/four-held-for-impersonating-jmpd-1.1823889#.VdrhYbKqqko>
- [14] Born, Six arrested for impersonating police officers in Pretoria. Times Live, 28 July, 2015. <http://www.timeslive.co.za/local/2015/07/28/Six-arrested-for-impersonating-police-officers-in-Pretoria>
- [15] Shain Germaner, Cloned cop car recovered in Gauteng. Iolnews, August 9 2013, <http://www.iol.co.za/news/crime-courts/cloned-cop-car-recovered-in-gauteng-.1560092#.VdrrmPbKqqko>
- [16] N. Carpenter and N. Glover, Radio Frequency Identification: Technology in the Federal Government. Report of Congressional Requesters, US Government Accountability Office, 2005
- [17] Linda C. and Samuel F.W. An Inside Look at RFID Technology. Journal of Tech. Management and Innovation, Pp.128-114, Vol.2, No. 001, March 2007
- [18] Lewis S. Basic Introduction to RFID and its Uses in the Chain Supply, LARAN RFID, TECH white paper publication, May 2004.
- [19] Patnaik S. RFID Solution for Assets Tracking and Inventory Management, FicuSoft, TECH white paper publication, 2004
- [20] Robert, C. M. Radio Frequency Identification (RFID). Computers & Security, Vol. 25. Pp. 18 – 26, 2006
- [21] Bakery, N.S. et al. RFID Application in Farming Management System. In Proceeding of 3rd International Conference on Robotics, Vision, Information and Signal Processing 2007 (ROVIS2007), Penang, 28 – 30 November 2007
- [22] Yan, B. and Lee, D. Application of RFID in Cold Chain Temperature Monitoring System. 2009 ISECS International Colloquium on Computing, Communication, Control, and Management. Aug. 8 – 9, 2009. Sanya, China
- [23] Voulodimos, A. S. et al. A Complete Farm Management System based on Animal Identification using RFID Technology. Computers and Electronics in Agriculture. Vol. 70. Pp. 380 – 388, 2010.
- [24] Haron, N. S., Saleem, N. S., Hassan, M. H., Ariffin, M. M. and Aziz, I. A. A RID-based Campus Context-Aware Notification System. Journal of Computing. Vol. 2. Issue 3, 2010
- [25] Herdawatie et al. Fusion of Radio Frequency Identification (RFID) and Fingerprint in Boarding School Monitoring System (BoSs), Sustainable Radio Frequency Identification Solutions, Cristina Turcu (Ed.), InTech, 2010
- [26] Christopher L. Howertona, Joseph P. Garner b, Joy A. Mencha. A system utilizing radio frequency identification (RFID) technology to monitor individual rodent behavior in complex social settings. Journal of Neuroscience Methods 209 (2012) 74–78
- [27] Rama, N.S., Prasad, K. and Rajesh, A. RFID-Based Hospital Real Time Patient Management System. International Journal of Computer Trends and Technology- volume 3, Issue 3- pp.509, 2012 ISSN: 2231-2803
- [28] Catarinucci, L. et al. An animal tracking system for behavior analysis using radio frequency identification. Lab Anim (NY). 2014 Sep, 43(9):321-7. doi: 10.1038/labani.547
- [29] Auto-ID Center. Technology Guide, Auto-ID Center, 2002, [www.autoidcenter.org](http://www.autoidcenter.org), 27/08/2015
- [30] Wikipedia. EPCglobal Network. [https://en.wikipedia.org/wiki/EPCglobal\\_Network](https://en.wikipedia.org/wiki/EPCglobal_Network). 02/09/2015
- [31] Wikipedia. Auto-ID Labs [https://en.wikipedia.org/wiki/Auto-ID\\_Labs](https://en.wikipedia.org/wiki/Auto-ID_Labs). 02/09/2015

# AN (M, K) MODEL BASED REAL-TIME SCHEDULING TECHNIQUE FOR SECURITY ENHANCEMENT

Chandra Mouli, Smriti Agrawal

Department of Information Technology, Chaitanya Bharathi Institute of Technology, Hyderabad, India

**Abstract**—Real Time Systems are systems where timely completion of a task is required to avoid catastrophic losses. The timely completion is guaranteed by a scheduler. The conventional schedulers only consider only the meeting the deadline as only design parameter and do not consider security requirement of an application. Thus modification of the conventional scheduler is required to ensure security to the Real time application. The existing security aware Real-Time scheduler (MAM) provides security to a task whenever possible and drops tasks whenever it is unable to schedule it within its deadline. The major problem in this tech scheduler is that it does not guarantee minimum security to all tasks. Thus, some may be exposed to security threats, which may be undesirable. Further, tasks are dropped in an unpredicted manner which is undesirable for Real Time Systems. This project presents an (M, K) model based Real Time scheduling technique SMK which guarantees that ‘M’ tasks in a window of ‘K’ successive task complete. It guarantees minimum security level to all the tasks and improves whenever possible. The simulation results show that the proposed SMK is approximately 15% better than the existing system.

**Keywords**—Hard real-time scheduler, security, (M, K) model.

## I. INTRODUCTION

Real-time systems are those systems that have stringent constraints[1]. Real time systems can be classified Hard, Soft and Weakly hard real-time systems. In hard real time systems [2], time constraints must always be honored. That means the tasks should complete their execution before their deadline. Some of the examples for hard real-time systems are traffic control, medical emergencies, and aircraft control etc. In soft real-time systems[3], time constraints are considered but quality of the output decreases with delay in the output. Some of the examples of soft real-time systems are online transactions, buses arrival time, etc. However, majority of applications are weakly hard real-time systems [4] in which every accepted job of a task must meet its deadline while some may be compromised. For real-time data, quality of service is applied for real time audio and video streams without congesting the network. To provide security for different services real time data packets sources are required[5],[6]. To maintain the balance between packet delivery and security effectively. A model known as (M, K) model is used for the selection of the packet in real-time scheduling for providing security. This project has scheduler which is used to schedule the tasks that are given as input and it schedules the tasks according to shortest deadline. For that we are giving security for those tasks/jobs according to their execution time. If there

is a scope of increasing its security level then its value should be added to execution time and it should not exceed the deadline. If that condition satisfies then the additional security is given for that task/packet. This paper is organized as follows. Section II is Literature Survey in this previous papers related to real-time scheduling is been discussed, Section III is System Design in this the system architecture is been discussed, Section IV is Implementation in this the algorithm and the way it goes is been presented, Section V is Simulation Results in this the results of proposed work is been discussed, Section VI is Conclusion in this the summary of the proposed work is been discussed and Section VII is References in this related papers are been kept.

## II. LITERATURE SURVEY

### 2.1 (M, K) MODEL

Here this model is used to decide which security module is taken and how much security level should be increased. This decision of increasing the security level is based on different methods which are as follows:

1. Deeply Red\_Pattern(Red\_Pattern): This pattern was proposed by authors [7]. Mathematically represented as

$$\pi_i^j = \begin{cases} 1, & \text{if } j = \left\lfloor \left[ \frac{j * m_i}{k_i} \right] * \frac{k_i}{m_i} \right\rfloor \text{ for } j = 0, 1, \dots, k_{i-1} \\ 0, & \text{otherwise} \end{cases}$$

Here, release mandatory while it is optional in case 0 is assigned to. We refer this pattern as Red\_Pattern. Advantage of applying this pattern to a task set for security decision is that it aligns the optional jobs together so that a component has a better opportunity to switch into sleep state to save energy. For a task whose critical speed is higher than or equal to the highest possible speed the operating speed should never be scaled down. Assigning Red\_Pattern to such a task helps to extend the idle interval for switching to sleep state. However, for a task whose critical speed is lower than Red\_Pattern overloads the system leading to large size busy intervals and need more energy to be feasible.

### 2.2 REAL-TIME MULTI-AGENT DESIGN MODEL

#### A. Source Agent

The source agent generates the data packets. It will generate either audio or video type real-time data packet. Each packet has a fixed packet size i.e.  $p_s=1500$ , which is the maximum



frame length of Ethernet frame payload. The real time traffic is sent by the source agent with the rate of  $\lambda_f$ . For modeling the packet inter-arrival time an exponential distribution with mean  $1/\lambda_f$  is used.

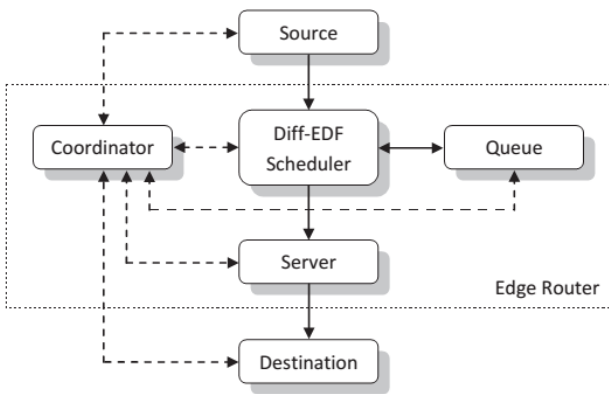


Fig.1. Multi-agent Model system for Real time network [9].

### B. Coordinator Agent

The coordinator agent acts as a software agent. To regulate their functionalities it interacts with other agents. It does not have entire view of the overall system. Using the known IP address it interacts with the source agent and using known MAC address it interacts with the destination agent to locate at the edge router.

### C. Diff-EDF Scheduler Agent

To provide the QoS requested by the source the Diff-EDF scheduler agent enforces the timing constraints on the packets. The Diff-EDF is based on the EDF scheduling algorithm and it is one of the algorithms that have real-time priority scheduling. While compared to other real-time schedulers such as EDF and FCFS the Diff-EDF scheduler shows smaller miss ratio and shorter average total packet delay at the edge router [8]. The Diff-EDF uses the effective deadline  $Def$  as the priority key instead of using the relative deadline.

### D. Server Agent

Scheduler chooses the real-time data packets and it is the responsibility of the server agent to serve them. The packet's remaining time till expiration will be the basis for serving or dropping of the packet by the server agent. If the packet does not expire, it will send it to the destination through their MAC address and it specifies service time with exponentially distributed. For packet server time an exponential distribution with mean  $1/\mu_f$  is used, where  $\mu_f$  is defined as packet service rate given by

$$\mu_f = B_w / (8P_s) \quad (1)$$

Where  $B_w$  is the average aggregate bandwidth needed for both types of real-time traffics (audio and video). the server is responsible for recording and keep tracking of a QoS parameter, the miss ratio, and reports it to the coordinator. In order to meet the QoS requirements the coordinator then adjusts system parameters accordingly.

### E. Buffer Queue Agent

The buffer queue agent has two processes: the queuing (storing) process and the dequeuing (fetching) process. In the queuing process, according to their effective deadlines the

queue agent will place the arriving packets in its buffer. When scheduler requests something this will be the response from the queue agent. In the dequeuing process, the queue agent searches for the packet which is going to expire and sends that packet to the scheduler. Then the same packet is forwarded to the server by the scheduler. The queue notifies/sends message of its buffer usage through a direct link which is established between queue agent and coordinator.

### F. Destination Agent

When a packet is received from the server the destination agent performs FCFS scheduling algorithm on it. It sends two parameters to the coordinator. The first it sends its processing rate  $P_f$  of traffic flow. At the initiation phase only it is sent to the coordinator. The second is it sends the size  $B_f$  of the available buffer for accommodating the packets of traffic flow. A time period  $t$  is specified by the coordinator. The destination agent sends  $B_f$  to the coordinator at the end of every time period. To determine the packet's security service level this process is used.

## 2.3 SECURITY SERVICE DESIGN [9]

Security services are applied to the data packets by the packet generators (sources) to combat network security threats. The proposed system i.e. multi-agent system will provide adaptive security enhancement rather than providing new security measures of the real-time data packets. As the LAN environment is the most vulnerable, it makes them robust against different security attacks/threats.

In this system we have two modules, namely the single-layer security service and the weighted multi-layer security service. Any module chosen can depend upon following measures the types of security threats, the processing capabilities of the end users, the NPMs, and the QoS requirements.

### A. Single-Layer Security Service

For the single-layer security service, the system will provide one of the following security services for the real-time data packets: confidentiality (c), integrity (g), or authentication (a).

TABLE I: CONFIDENTIALITY ALGORITHMS

Index(j)	Algorithm	$S'_j$	$\mu'_j$ (KB/ms)
1	SEAL	0.08	168.75
2	RC4	0.14	96.43
3	Blowfish	0.36	37.5
4	Knufu/Khafre	0.40	33.75
5	RC5	0.46	29.35
6	Rjindael	0.64	21.09
7	DES	0.90	15
8	IDEA	1.00	13.5

For confidentiality security service, to each real-time data packet a security level ranging from 1 to 8 is assigned. In that one of the eight security levels are used by the source agent as cryptographic algorithms. In that index 1 indicates that it is the weakest algorithm and in same manner index 8 indicates that it is the strongest algorithm. The experiments performed on a 175-MHz-processor machine are responsible to show the confidentiality security algorithms and shows the processing rates in table 1.

$$S_j^c = \mu^c / \mu^j = 13.5 / \mu^j \quad (2)$$

TABLE II: INTEGRITY ALGORITHMS

Index(j)	Algorithm	$S_j^c$	$\mu_j^c$ (KB/ms)
1	MD4	0.18	46.4
2	MD5	0.26	33.2
3	RIPEMD	0.36	23.3
4	RIPEMD-128	0.45	18.9
5	SHA-1	0.63	13.4
6	RIPEMD-180	0.77	11.1
7	Tiger	1.00	8.5

For the integrity security service, to protect the real-time data packet from the alteration security threat different hash functions are implemented. The experiments performed on a 175-MHz-processor machine are responsible to show the integrity security algorithms and associated with the processing rates in table II at the source agents. In the table, the packet ranges from 1 to 7 and the integrity efficiency factor,  $S_j^c$  ranges from 0.18 to 1. the efficiency factor is given by

$$S_j^c = \mu_j^c / \mu_j = 8.5 / \mu_j \quad (3)$$

TABLE III: AUTHENTICATION ALGORITHMS

Index(j)	Algorithm	$S_j^a$	$\mu_j^a$ (KB/ms)
1	HMAC-MD5	0.55	16.1
2	HMAC-SHA-1	0.91	9.1
3	CBC-MAC-AES	1	8.8

For the authentication security service, different security encryption algorithms are implemented. The experiments performed on a 175-MHz-processor machine are responsible to show the authentication security algorithms and associated with the processing rates at the source agents. In the table, the packet ranges from 1 to 3 and the authentication efficiency factor,  $S_j^a$  ranges from 0.55 to 1. the efficiency factor is given by

$$S_j^a = \mu_j^a / \mu_j = 8.8 / \mu_j \quad (4)$$

#### B. Weighted Multi-Layer Security Service

With the Weighted Multi-Layer Security Service, according to the feedback from the coordinator agent on the real-time data packet the source agent applies the optimal algorithm for each security service (confidentiality, integrity, and authentication). To anyone of the security algorithms responds to the destination agent for each processor. Therefore, we overcome the overhead solving the three security algorithms sequentially.

Based on the estimation of the resources (available memory and processing speed) of the multi-processor destination agent the proposed system needs to adaptively enhance the packet security levels. This can be answered by predefining a security threshold serving vector,  $\psi$ , at network initiation. Based on the security requirements defines the weights of the security services by the source agent.  $\psi = (\psi^c, \psi^g, \psi^a)$ , where each element of the vector reflects the weight for each security service (confidentiality, integrity, and authentication, respectively). The destination nodes shared memory portion is reserved by each processor accordingly, so that

$$\sum_{i=\{c,g,a\}} \psi_i = 1 \quad (5)$$

#### 2.4 SECURE DIFF-EDF SCHEDULER MULTI-AGENT SYSTEM

Based on the conventional static queuing theory the proposed real-time scheduling with security awareness cannot be modeled due to its adaptive feature. The coordinator models the scheduling process as a general Brownian motion with a negative motion drift parameter  $(-\theta)$ , upon receiving a request from the source [10].

$$\theta = \frac{2(1-I)}{\sum_{f=1}^N (I_f^2 \sigma_{1f}^2 + \sigma_{2f}^2)} \quad (6)$$

Where  $\sigma_{1f}$  is the standard deviation of the inter-arrival time for real-time traffic flow  $f$ ,  $\sigma_{2f}$  is the standard deviation of the service time for flow  $f$ .

$I_f$  is the intensity of traffic flow  $f$  that is given by

$$I_f = \lambda_f / \mu_f \quad (7)$$

Where  $\mu_f$  is the packet service rate and  $\lambda_f$  is the packet sending rate for traffic  $f$ . With  $N$  real-time traffic flows, the total intensity of all flows  $I$  is given by

$$I = \sum_{f=1}^N I_f \quad (8)$$

$\Phi_{min}$  be the smallest deadline miss ratio of all flows. The coordinator calculates the parameter  $C_f$  as

$$C_f = \theta^{-1} \log(\Phi_f / \Phi_{min}) \quad (9)$$

Where  $\Phi_f$  is the required deadline miss ratio of traffic  $f$ . By estimating the flow deadline miss ratio, the coordinator evaluates the feasibility of serving such request

$$\hat{\Phi} = \exp(-\theta(D_{avg} - C_f)) \quad (10)$$

Where the average effective deadline for all data flows is  $D_{avg}$ .

$$I = \sum_{n=1}^{\infty} I_f (\Phi_f + C_f) \quad (11)$$

The coordinator performs the following actions: (1) Sending an acceptance message to the source; (2) Passing the parameter  $C_f$  to the Diff-EDF scheduler; and (3) Passing the required deadline miss ratio  $\Phi_f$  to the server, if the estimated deadline miss ratio meets the QoS requirement, i.e.  $\Phi_f < \Phi_f$ . By sending its real-time data packets to the scheduler source agent responds to the acceptance message. To obtain the effective deadline of the packets the scheduler performs a shadow function.

$$D_{ef} = D_f + C_f \quad (12)$$

The queue agent will queue the packets based on their effective deadlines when the scheduler forwards the packets.

The queue agent fetches a packet that is closest to expire (with smallest  $D_{ef}$ ) and forwards it to the scheduler. The scheduler passes the packet to the server. Once it receives the packet, the server agent performs the following:

- (1) Changing its current status to busy;
- (2) Serving the unexpired packet (deadline is not exceeded) or dropping the expired packet;
- (3) Forwarding the packet to its destination according to its MAC address; and
- (4) Keeping track of two counters:  $n_f$  the number of served packets of traffic flow  $f$  and  $t_f$  the sum of time differences of packets of traffic  $f$  arrived at its destination, that is

$$t_f = \sum_{i=2}^{n_f} (t_{i,f} - t_{i-1,f} = t_{n_f,f} - t_{1,f}) \quad (13)$$

The coordinator interacts with the server and the destination requesting for the server's counter information ( $n_f$  and  $t_f$ ) and the destination's resource information (processing rate ( $P_f$ ) and size of available buffer ( $B_f$ )) for every time period  $T$ . After receiving any information from the coordinator finds the mean inter-arrival time,  $1/\zeta_f$ , for the packets of traffic flow delivered to their destination. We have

$$1/\zeta_f = t_f / (n_f - 1) \quad (14)$$

The coordinator interacts with the server when it notices a miss ratio near the set miss ratio limit  $\Phi_f$ . The source adjust its parameters such as reducing the sending rate or raising the deadline miss ratio limit if the coordinator suggest it to do. After that it does some changes and updates the same to scheduler  $C_f$  and server  $\Phi_f$ .

The above interactions among the agents are common for both single-layer and weighted multi-layer security service modules. The module which we are using decides the security enhancement process performed by the coordinator agent.

#### A. Security Service with Single-Layer Module

Different processing rates  $\{\mu\}$  are stored by coordinator as shown in table I, II, and table III. Using the  $j$ th security algorithm of then security service, it determines the length of the buffer,  $L_{fj}^x$ , that is needed to enhance  $n_f$  real-time packets and  $x$  indicates  $c$  for confidentiality,  $g$  for integrity,  $a$  for authentication service. We have  $L_{fj}^x = P_f^x \zeta_f$  (15)

$P_f^x$  is the total processing time for the packets of traffic flow  $f$ . It takes into account two delays: the delay of resolving the conflict of two or more equally prioritized packets,  $D_{f, equal\_priority}$  and the delay due to the preemption process when the arrived packet is closer to expire than the remaining time of the current packet processing,  $D_{f, preemption}$ . Therefore,

$$P_f^x = D_{f, equal\_priority} + D_{f, preemption} + J_{fj}^x \quad (16)$$

Where  $J_{fj}^x$  is the time required to process a packet of length  $P_s = 1.46KB$  (1500 bytes) using the  $j$ th security algorithm of the  $x$  security service. It is given by

$$J_{fj}^x = \frac{P_f}{\mu_j^x \beta_f} = \frac{1.46KB}{\mu_j^x \beta_f} \quad (17)$$

As defined before  $P_f$  is the processing rate of traffic flow  $f$  at the destination agent. To destination machines we can apply the proposed scheme with different processor speed other than

175 MHz research has been done on this by taking values as 266MHz and 2.4 GHz and results evaluate that the performance of security algorithms in processing rate can be linearly related to the processor speed.

The length of available buffer at the destination of traffic  $f$  is  $B_f$  and  $L_{fj}^x$  is the length of buffer needed to enhance  $n_f$  packets to security level  $z$  of the  $x$  security service, the coordinator enhances/reduces security to level  $z$  or maintains the same level such that  $L_{fz}^x \leq B_f < L_{f(z+1)}^x$  (18)

The coordinator notifies the source, once the decision on security level is made. If the decision is to stay at the current security level then no notification is sent. The source agent applies corresponding new security enhancement algorithm to the packets to be sent only after receiving notification.

#### B. Security Service with Weighted Multi-Layer Module

The weighted multi-layer module applies multiple security enhancements security service to the real-time packets, the coordinator agent designs the best security level for each security service to be adopted by the source. The coordinator depends on two factors to perform its security enhancement algorithm: the congestion control feedbacks and the pre-determined weights of the security services.

The source agents specifies security requirements in the threshold serving vector  $\psi_f = (\psi_f^c, \psi_f^g, \psi_f^a)$ , the portion of the destination's available buffer is evaluated by the coordinator,  $B_f^x$ , where  $x$  indicates  $c$  for confidentiality security service,  $g$  for integrity security service, or  $a$  for authentication security service. We have  $B_f^x = \psi_f^x B_f$  (19) And at the destination overall available buffer  $B_f$  is given by

$$B_f = \sum_{i=\{c,g,a\}} B_f^i \quad (20)$$

To enhance  $n_f$  packets using different security algorithms with the required length of buffer the coordinator compares the length of available buffer at the destination. The coordinator enhances/reduces security level or maintains the same level such that  $L_{fz}^x \leq \psi_f^x B_f < L_{f(z+1)}^x$  (21)

### III. SYSTEM DESIGN

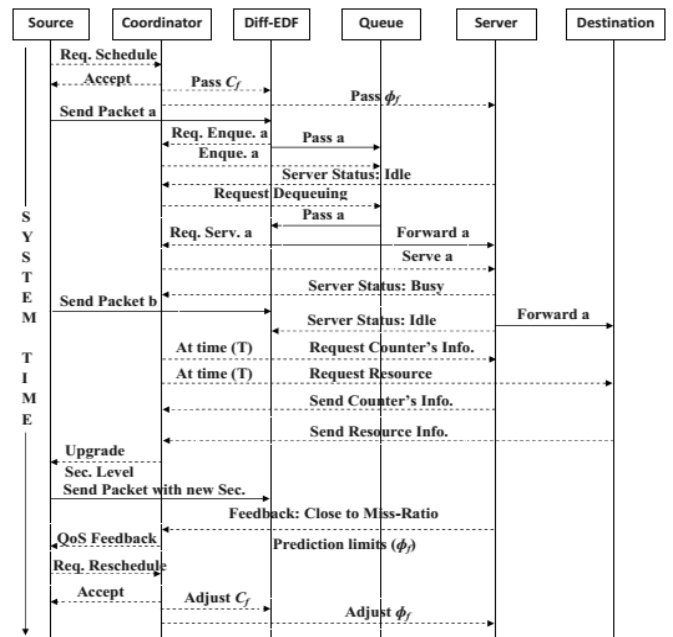


Fig. 2. Timing diagram of interactions and data transfers in the multi-agent system.

**A. Source:**

Data packets are been generated in this module. The real –time data packet generated is either audio or video. The packet has a fixed packet size i.e.  $p_s=1500$ , which is equal to the maximum frame length of Ethernet frame payload. The packet generated by the source is send to the Packet Coordinator.

**B. packet Coordinator:**

The Packet coordinator agent acts as a software agent. To meet the other agent’s functionalities it interacts with other agents. It will not know the systems entire view.Using the known IP address it interacts with the source agent and using known MAC address it interacts with the destination agent to locate at the edge router. The Packet coordinator will send the packet to real-time scheduler to check whether the packet is feasible or not.

**C. Real time scheduler and security enhancer:** Here EDF is taken as the scheduler and it takes the shortest deadline as the priority and sets that task in the first order. Every time a new packet arrives the queue is rearranged. Here EDF is used as the real-time scheduler agent. If the packet is not feasible it is rejected and the information is passed to the Packet Coordinator. After scheduling is completed security level is increased for the packet and checked if its deadline is not effected after increasing the security. If the security level increased and if it is not affecting the deadline of the packet then its security level is increased successfully.

**D. Network Queue:** The Network queue has two processes: the queuing (storing) process and the dequeuing (fetching) process. In the queuing process, according to their effective deadlines the queue will place the arriving packets in its buffer. When scheduler requests the packet it will give the response by sending the requested packet. In the dequeuing process, the queue agent searches for the packet which is going to expire and sends that packet to the scheduler. Then the same packet is forwarded to the server by the scheduler. The network queue notifies/sends message of its buffer usage through a direct link which is established between queue agent and coordinator. Here the secured packet generated by the real-time scheduler agent is been sent to the destination.

**E. Destination:** When a packet is received from the Network Queue the destination agent performs FCFS scheduling algorithm on it. The acknowledgment is been given to the Packet Coordinator by the destination agent whether the packet is been received or not successfully.It sends twoparametersto the coordinator. The first it send its processing rate  $P_f$  of traffic flow  $f$ . At the initiation phase only it is send to the coordinator. The second is it sends the size  $B_f$  of the available buffer for accommodating the packets of traffic flow  $f$ . A time period  $t$  is specified by the coordinator. The destination agent sends  $B_f$  to the coordinator at the end of every time period. To determine the packet’s security service level this process is used.

**Flow chart:**

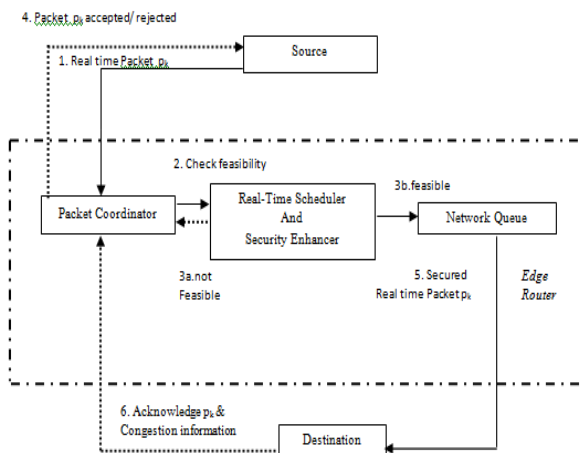
Figure 3.2 describes how the flow is organized from the packet which is ready to schedule. From the admission queue the packet is sent to feasibility test. Before that the network is checked whether it is congested or not. If the network is congested then the packet is sent for feasibility test and if it is feasible, then it is sent to the dispatch queue and if the packet is not feasible then it is rejected.

The packets which are sent in this manner are considered as optional packets. If congestion is not their then the packet is checked whether it is optional or not. If the packet is optional then the packet is been rejected other wise the packet is accepted and sent for the feasibility test. After the feasibility test is completed the security level of that packet is increased.

The feasibility test is done whether the packet has the minimum security and whether the packet will be executed successfully for sending to destination. In this the packet is first estimated for the finish time of that packet. If the finish time which is estimated is not more than the deadline then that packet is accepted. And if that packet is kept in the queue and if another packet comes and if that deadline is nearer than the others then it is tested for feasibility test and again the order of that queue is arranged once again.

The packet will be processed and that packet will complete its execution within the deadline. Then the security comes. Here the security is increased if only the enhancement is fruitful. The packet is given more security if it is mandatory packet and if its security is added also and the packets deadline is not missed. Because if the packets deadline is missed then there is no point in increasing the security. So the feasibility test should be fruitful.

After increasing the packet security level the packet should not miss the deadline. The packet is sent to dispatch queue. It is decided by (M, K) model. This has a pattern named Deeply Red\_Pattern (Red\_Pattern).if the value is shown as 1 the security level of that packet is increased otherwise it is shown as 0 and the security level is not increased. If the value is 1 then it is mandatory and if it is 0 then it is considered as optional. The optional packets which are sent also have minimum security level.



**Fig 3: Flow of the Proposed System**

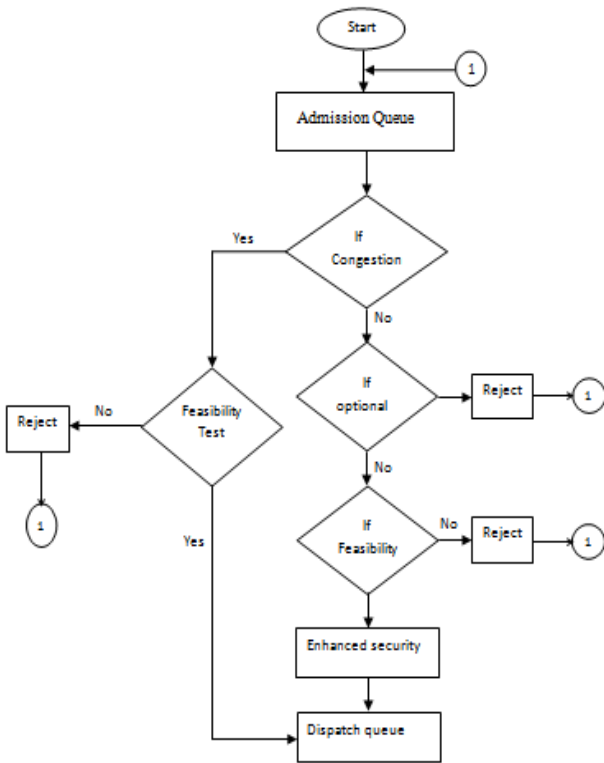


Fig 4: Packet flow in the proposed SMK

**Algorithm security enhancement (task  $(\tau_i^j)$ )**

Begin

1. while (admission queue is empty) do

1.1 wait

end while

2. Select job  $\tau_i^j$  from admission queue

3. estimate  $ft_i^j$  with minimum security level

4. if (congestion) then

begin if

4.1 if (not feasible( $\tau_i^j$ ))

4.1.1 reject  $\tau_i^j$

else

4.1.3 Send the job to dispatch queue

4.2 goto step1

else

4.3 if ( $\tau_i^j$  is optional)

4.3.1 reject  $\tau_i^j$

4.3.2 goto step 1

4.4 if (feasible ( $\tau_i^j$ ))

4.4.1 enhance security

4.4.2 send to dispatch queue

else

4.4.3 reject  $\tau_i^j$

end

4.5 goto step1

end

**//Function to perform feasibility test//**

**Feasible (job  $\tau_i^j$ )**

1. estimate

$$ft_i^j = a_i^j + wt_i^j + e_i^j \text{ where } wt_i^j = \max(ft_h^k) - a_i^j$$

//  $\tau_h^k$  is a higher priority job in the dispatch queue//

2. if ( $ft_i^j > D_i^j$ ) //will miss its deadline

```

2.1 reject  $\tau_i^j$  and goto step 1
else
//check if the new job will force the lower priority jobs already
accepted to miss their deadlines//lower priority job
begin
2.2 for every job  $\tau_i^k$  in the dispatch queue
// $\tau_i^k$  lower priority job//
do
begin for
2.2.1  $ft_i^k = ft_i^k + e_i^j$ 
//reestimate the finish time of a lower priority job//
2.2.2 if ( $ft_i^k > D_i^k$ ) //will miss its deadline
begin
2.2.2.1 reject  $\tau_i^j$ 
2.2.2.2  $ft_i^k = ft_i^k - e_i^j$ 
2.2.2.3 goto step 1
end
end for
end
end

```

The algorithm explains how the scheduling is been done using EDF. For that tasks security is provided where ever it is required. The detailed implementation of project is as follows; For task set  $T_i = (a_i, e_i, f_i, d_i, l_i, s_i)$  are taken.

$a_i$ =arrival time

$e_i$ =execution time

$f_i$ =duration time

$d_i$ =deadline

$l_i$ =amount of data (measured in KBs)

$s_i$ =security level

The source generates packets which are sent to the admission queue and its deadline is estimated. It is then treated as a real time task. When the tasks are entered into the admission queue. If the admission queue is not empty then we should wait for sometime. Then a job  $\tau_i^j$  is selected from the admission queue. After that the finish time of that job is to be estimated with minimum security level. After that we should check whether the network is congested or not. If the network is congested then the optional packets are to be sent through network. but here the job is checked for feasibility.

If the job is feasible then the job is sent to dispatch queue with minimum security otherwise it is rejected. If the network is not congested then the job is checked whether it is optional packet or mandatory packet. If the packet is optional then that packet is rejected otherwise that packet is accepted. After that packet is accepted then that packet is sent for feasibility test. If the packet is feasible then its security level is been increased and sent to dispatch queue. If the packet is not feasible then that packet is rejected.

Here the feasibility test is done in a process which is as follows.

For a job  $\tau_i^j$  we have to estimate finish time as  $ft_i^j = a_i^j + wt_i^j + e_i^j$  where waiting time of the job would be  $wt_i^j = \max(ft_h^k) - a_i^j$ , if the  $\tau_h^k$  is a higher priority job in the dispatch queue. after that we have to look whether the finish time is greater than the deadline. if the finish time is greater then that job would be rejected. Because this job can disturb the already accepted jobs. The above is about in the case if it is a higher priority job. Now if the lower priority job comes then again we have to reestimate the finish time of that job.

Now we have to check whether it is missing its deadline. If it is missing its deadline then that job will be rejected.

Coming to security part the user is provided with 3 security services and he is free to choose any of the service. The services provided to the user are confidentiality, integrity and authentication. In that there are encryptions algorithms specified for each security service separately.

When the user selects any one of the security service the algorithms present in that service are mentioned and in that user chooses any of the algorithm. Now the security enhancement is given by (m, k) model. Here in this (m, k) model there are some patterns which are as follows deeply red pattern, evenly distributed pattern, reverse pattern, hybrid pattern, mixed pattern.

From the above patterns in this project the author has taken deeply red pattern and implemented for the enhancement of the tasks. In this deeply red pattern we take the threshold value according to the execution time and if it is above the threshold value its value is 1 and security is increased. Otherwise security is not increased. The tasks for which security is increased its deadline should not be missed. If the packets deadline is missed, then there is no point in increasing the security for particular task.

The author tries to increase the security level of the packet according to the execution time. If the execution time is less than the deadline then the security level is increased.

#### 4.1 SECURITY REQUIREMENTS [11]:

A task consists of set such as  $T_i = (a_i, e_i, f_i, d_i, l_i, s_i)$  Where  $a_i$ ,  $e_i$ , and  $f_i$  are the arrival, execution, and finish times,  $d_i$  is the deadline, and  $l_i$  denotes the amount of data (measured in KB) to be protected.  $e_i$  can be estimated by code profiling and statistical prediction [12]. Suppose  $T_i$  requires  $q$  security services represented by a vector of security level ranges, e.g.

The vector characterizes the security requirements of the task.  $S_i^j$  is the security level range of the  $j$ th security service required by  $T_i$ . The security level determines the most appropriate point  $s_i$  in space  $S_i$ ,

e.g.  $s_i = (s_i^1, s_i^2, \dots, s_i^q)$  where  $s_i^j \in S_i^j, 1 \leq j \leq q$

It is imperative for a security-aware scheduler to adopt away of measuring security benefits gained by each admitted task. As such, the security benefit of task  $T_i$  is quantitatively modeled as a security level function denoted by  $SL: S_i \rightarrow R$

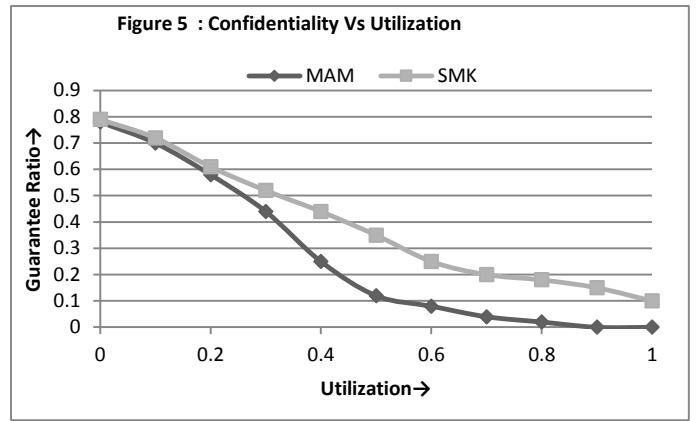
$$SL(S_i) = \sum_{j=1}^q W_i^j s_i^j, 0 \leq W_i^j \leq 1, \sum_{j=1}^q W_i^j = 1 \quad (22)$$

$W_i^j$  is the weight of the  $j$ th security service for task  $T_i$ . Users specify in their requests the weights to reflect relative priorities of the required security services.

$X_i$  denotes all possible schedules for task  $T_i$  and  $x_i \in X_i$  is a scheduling decision of  $T_i$ .  $x_i$  is a feasible schedule if 1) deadline  $d_i$  can be guaranteed, i.e.,  $f_i \leq d_i$ , and 2) the security requirements are met, i.e.  $\min(S_i^j) \leq s_i^j \leq \max(S_i^j)$ .

Given a real-time task  $T_i$ , the security benefit of  $T_i$  is expected to be maximized by the security controller under the timing constraint:  $SB(X_i) = \max_{x_i \in X_i} \{SL(s_i(x_i))\}$

$$= \max_{x_i \in X_i} \left\{ \sum_{j=1}^q W_i^j s_i^j(x_i) \right\} \quad (23)$$



Where the security level of the  $j$ th services  $s_i^j(x_i)$  is obtained under schedule  $x_i$ , and  $\min(S_i^j) \leq s_i^j(x_i) \leq \max(S_i^j)$ .  $\min(S_i^j)$  and  $\max(S_i^j)$  are the minimum and maximum security requirements of task  $T_i$ .

A security-aware scheduler aims at maximizing the system's quality of security, or security value, defined by the sum of the security levels of admitted tasks (see (22)). Thus, the following security value function needs to be maximized, subject to certain timing and security constraints:

$$SV(X) = \max_{x \in X} \left\{ \sum_{i=1}^p y_i SB(x_i) \right\} \quad (24)$$

Where  $p$  is the number of submitted tasks,  $y_i$  is set to 1 in task  $T_i$  is accepted, and is set to 0 otherwise. Substituting (23) into (24) yields the following security value objective function. Our proposed security-aware scheduling algorithm strives to schedule tasks in a way to maximize (25):

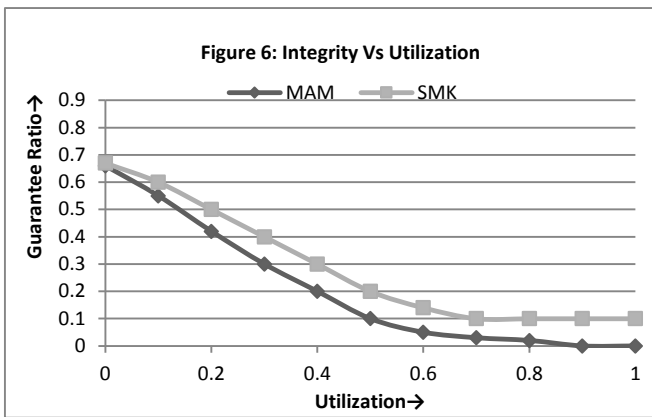
$$SV(X) = \max_{x \in X} \left\{ \sum_{i=1}^p \left( y_i \max_{x \in X} \left\{ \sum_{i=1}^p y_i SB(x_i) \right\} \right) \right\} \quad (25)$$

## IV. SIMULATION RESULTS

The effect of the variation of the Confidentiality is shown in the figure 5.1. The graph plotted has guarantee ratio from (0-0.9) and utilization from (0-1). When the utilization is increased (0-0.3) it can be observed that the guarantee ratio will be decreased in MAM (Multi Agent Model) gradually compared to SMK (Secure MK) and the values which it meets lower than that of the SMK approach, while this reduction in the guarantee ratio is more gradually decreased at higher values of the utilization (0.4-1). The MAM has less efficiency than the SMK and the algorithms which used are same but the schedulers are different. The difference shows that there is an increase in guarantee ratio by 15%.

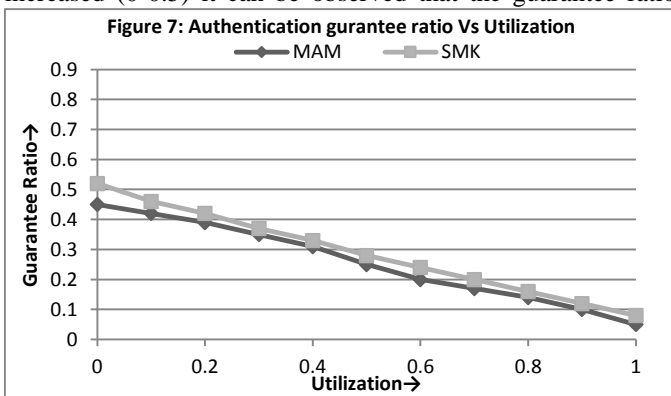
The effect of the variation of the Integrity is shown in the figure 5.2. The graph plotted has guarantee ratio from (0-0.8) and utilization from (0-1). When the utilization is increased (0-0.5) it can be observed that the guarantee ratio will be decreased in MAM gradually compared to SMK and the values which it meets lower than that of the SMK approach, while this reduction in the guarantee ratio is more gradually



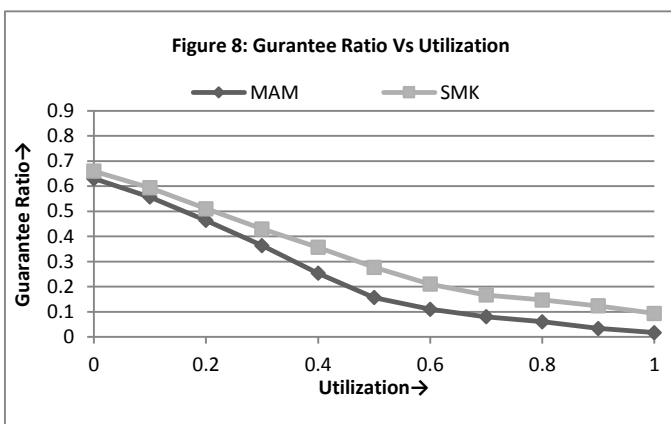


decreased at higher values of the utilization (0.5-1).The MAM has less efficiency than the SMK and the algorithms which used are same but the schedulers are different.The difference shows that there is an increase in guarantee ratio by 19%.

The effect of the variation of the Authentication is shown in the figure 5.3. The graph plotted has guarantee ratio from (0-0.6) and utilization from (0-1). When the utilization is increased (0-0.5) it can be observed that the guarantee ratio



will be decreased in MAM gradually compared to SMK and the values which it meets lower than that of the SMK approach, while this reduction in the guarantee ratio is more gradually decreased at higher values of the utilization (0.5-1).The MAM has less efficiency than the SMK and the algorithms which used are same but the schedulers are different. The difference shows that there is an increase in guarantee ratio by 12%.



The effect of the variation of the All Security Services is shown in the figure 5.1.The graph plotted has guarantee ratio from (0-0.7) and utilization from (0-1). When the utilization is increased (0-0.6) it can be observed that the guarantee ratio will be decreased in MAM (Multi Agent Model) more gradually compared to SMK(Secure MK) and the values which it meets lower than that of the SMK approach, while this reduction in the guarantee ratio is gradually decreased at higher values of the utilization (0.7-1).The MAM has less efficiency than the SMK and the algorithms which used are same but the schedulers are different.The difference shows that there is an increase in guarantee ratio by 15%.

## V. CONCLUSION

This work proposed an effective scheduling algorithm which provides improved security levels to the weakly hard real-time systems. It used (M, K) model where M packets are delivered with full security in K window of packets to provide derived QoS to the system. The existing systems provided best effort service, wherein it improved the security levels for some. However if the deadline did not permit the existing system passed the packet without any security. Further, based on the feedbackon packet loss the existing technique adjusted the security level provided to an individual packet to unpredictable packet loss. The proposed technique overcomes there limitations of the existing technique by providing minimum security to all the accepted packets. It partitions the packet as mandatory or optional using (M, K) model to provide better prediction on packet loss.

However,for any accepted packet if its deadline permits its security is also improved. The existing technique also does not take into account the packet losses due to congestion. The proposed algorithm also considered losses due to congestion in the network and proposed to send only mandatory packets to achieve the derived QoS in case there is no congestion in the network. However, if congestion exists and packets are dropped the proposed technique sends optional packets to provide best effort service. It is observed that when the system utilization is low (0-0.4) the proposed technique and existing technique provide similar security results. However, as the utilization ratio increased the proposed technique provides approximately 15% better security guarantee.

## REFERENCES

- [1]. Sha, L.et al.: Real Time Scheduling Theory: A Historical Perspective, Real-Time Systems 28,101-155 (2004).
- [2]. A. Burns, "Scheduling Hard Real-Time Systems: A Review," Software Engineering Journal, May 1991.
- [3].R. Al-Omari, A.K. Somani, and G. Manimaran, "An Adaptive Scheme for Fault-Tolerant Scheduling of Soft Real-Time Tasks in Multiprocessor Systems," *J. Parallel and Distributed Computing*, vol. 65, no. 5, pp. 595-608, May 2005.
- [4].G. Bernat, A. Burns, and A. Llamosi. Weakly Hard Real-time Systems. *IEEE Transactions on Computers*,vol. 50, no. 4, pp. 308 – 321, Apr. 2001.
- [5]. Y. Jung and M. Peradilla, "Tunnel gateway satisfying mobility and security requirements of mobile and IP-based networks," *J. Commun.and Networks*, vol. 13, no. 6, pp. 583–590, Dec. 2011.



- [6]. F. Hashim, K. S. Munasinghe, and A. Jamalipour, "Biologically inspired anomaly detection and security control frameworks for complex heterogeneous networks," *IEEE Trans. Network and Service Management*, vol. 7, no. 4, pp. 268–281, Dec. 2010.
- [7]. P. Ramanathan, "Overload management in realtime control applications using (m; k)-firm guarantee," *IEEE Trans. Parallel. Distrib. Syst.*, vol. 10, no. 6, pp. 549–559, Jun. 1999.
- [8]. M. Saleh and L. Dong, "Adaptive security-aware scheduling using multiagent system," in 2012 IEEE International Conference on Communications.
- [9]. MaenSaleh, and Liang Dong, "Real-Time Scheduling with Security Enhancement for Packet Switched Networks", *IEEE transactions on network and service management*, vol. 10, No. 3, September 2013.
- [10]. H. Zhu, J. P. Lehoczky, J. P. Hansen, and R. Rajkumar, "Diff-EDF: a simple mechanism for differentiated EDF service," in *Proc. 2005 IEEE Real-Time Technology and Applications Symposium*, pp. 268–277.
- [11]. T. Xie and X. Qin, "Scheduling Security-Critical Real-Time Applications on Clusters," *IEEE Trans. Computers*, vol. 55, no. 7, pp. 864–879, July 2006.
- [12]. T.D. Braun et al., "A Comparison Study of Static Mapping Heuristics for a Class of Meta-Tasks on Heterogeneous Computing Systems," *Proc. Workshop Heterogeneous Computing*, Apr. 1999.

# Comparing PSO and GA Optimizers in MLP to Predict Mobile Traffic Jam Times

W. Ojenge  
School of Computing  
TUK  
Nairobi, Kenya

W. Okelo-Odongo  
School of Computing  
UON  
Nairobi, Kenya

P. Ogao  
School of Computing  
TUK  
Nairobi, Kenya

**Abstract-** Freely-usable frequency spectrum is dwindling quickly in the face of increasingly greater demand. As mobile traffic overwhelm the frequency allocated to it, some frequency bands such as for terrestrial TV are insufficiently used. Yet the fixed spectrum allocation dictated by International Telecommunications Union disallows under-used frequency from being taken by those who need it more. This under-used frequency is, however, accessible for unlicensed exploitation using the Cognitive Radio. The cognitive radio would basically keep monitoring occupation of desirable frequencies by the licensed users and cause opportunistic utilization by unlicensed users when this opportunistic use cannot cause interference to the licensed users. In Kenyan situation, the most appropriate technique would be Overlay cognitive radio network. When the mobile traffic is modeled, it is easier to predict the exact jam times and plan ahead for emerging TV idle channels at the exact times. This paper attempts to explore the most optimal predictive algorithms using both literature review and experimental method. Literature on the following algorithms were reviewed; simple Multilayer perceptron, both simple and optimized versions of support vector machine, Naïve Bayes, decision trees and K-Nearest Neighbor. Although in only one occasion did the un-optimized multilayer perceptron out-perform the others, it still rallied well in the other occasions. There is, therefore, a high probability that optimizing the multilayer perceptron may enable it out-perform the other algorithms. Two effective optimization algorithms are used; genetic algorithm and particle swarm optimization. This paper describes the attempt to determine the performance of genetic-algorithm--optimized multilayer perceptron and particle-swarm-optimization-optimized multilayer perceptron in predicting mobile telephony jam times in a perennially-traffic jammed mobile cell. Our results indicate that particle-swarm-optimization-optimized multilayer perceptron is probably a better performer than most other algorithms.

Keywords – MLP; PSO; GA; Mobile traffic

## I. INTRODUCTION

Freely-usable frequency spectrum is dwindling quickly in the face of increasingly-greater demand [1]. The irony is that as mobile traffic the world over, especially, overwhelm the frequency allocated to it, some frequency bands such as for terrestrial TV are insufficiently used [2]. Yet the fixed spectrum allocation dictated by International Telecommunications Union (ITU) disallows under-used frequency from being taken by those who need it more. This under-used frequency is, however, accessible for unlicensed exploitation using the Cognitive Radio (CR). The cognitive radio would basically keep monitoring occupation of desirable frequencies by the licensed users and cause opportunistic utilization by unlicensed users when this opportunistic use cannot cause interference to the licensed users [3]. Some individuals state that the digital migration has vacated the bigger spectrum and mitigated the

shortage of spectrum. However, with the emergence of the phenomenon of Internet of Things (IoT), where 50 billion devices are forecast to demand communication frequency, the digital migration dividend is likely to expire quickly.

In Nairobi city, Kenya, where this study was taken, the four mobile service providers suffer grave shortage of spectrum. The Communication Authority of Kenya (CAK) publishes Quality of Service (QoS) reports every two years. The latest report on Quality of Service of mobile service providers is the 2012-2013 one [4]; Kenya's dominant service provider with 68% market share, Safaricom Ltd., registered the worst blocked calls rate of 11% against a target of <5%. According to [5] [6], GSM call blocking, which may be SD BLOCKING or TCH BLOCKING, can be caused by many reasons. It can be: Optimization issue such as improper definition of parameters, incorrect or inappropriate timer, un-optimized handover and power budget margins, and interference; or hardware problem such as faulty transmitters. However, the main cause is often congestion when there is no available channel for assignment of call. In 2012, Safaricom Ltd is on record for having requested CAK to allow it to use the analogue TV band of 694-790 MHz for deployment of broadband, a request rejected by CAK - with reasons - in their press release of [7]. In this country, digital TV channel occupation by licensed users is typically poor as there are only 5 licensed broadcasters. Many digital TV frequencies therefore lie idle at certain moments. The worst jam times in mobile telephony, within the most congested mobile cells, would need to be determined by the cognitive radio. The most appropriate technique would be Overlay cognitive radio network [8]. During such times, the cognitive radio would explore which TV channels are idle in order to cause the mobile service provider's base station controller (BSC) to opportunistically utilize those TV channels in the times that they are idle. When the mobile traffic is modeled, it is easier to predict the exact jam times and plan ahead for emerging TV idle channels at the exact times. There are existing studies which have attempted to predict jam times in mobile telephony traffic with varying degrees of accuracy as shall be outlined in the next section.

This paper attempts to explore the most appropriate modeling algorithm using both literature review and experimental method.

## II. RELATED STUDIES

The following are instances where mobile telephony traffic has been modeled and predicted. In [9], wavelet transformation least squares support vector machines is used to conduct busy telephone traffic prediction with impressive mean relative error of -0.0102. In [10], correlation analysis is firstly applied to the busy telephone traffic data to obtain the key factors which influence the busy telephone traffic. Then wavelet transform is used to decompose and reconstruct the telephone traffic data to get the low-frequency and high-frequency components. The low-frequency component is loaded into ARIMA model to predict, while the high-frequency component and the obtained key factors are loaded into PSO-LSSVM model to predict. Finally, a least error value of 1.14% is achieved by superposition of the predictive values. In [11], probabilistic predictive algorithms, Naïve Bayes, Bayesian Network and the C4.5 decision tree were used. Although churn is an attribute that is different from traffic levels, the predictive accuracy in the two attributes may not be distant from each other as prediction of churn uses a subset of the data used in prediction of traffic levels. Naïve Bayes and Bayesian Network perform better than C4.5 decision tree with Naïve Bayes performing best with a relative error of -0.0456.

The following are instances of comparison between the most popularly-used predictive algorithms.

In [12], the SVM of the Gaussian Radial basis Polynomial Function achieved an accuracy of 95.79% against 84.50% for the MLP. In [13], SVM of the radial basis function achieved an accuracy of 100% against the MLP with 95%. In [14], the SVM performed better than MLP by an average of 19% over the 2008 to 2012 period.

Although in all the above circumstances, the SVM is seen to out-perform the MLP, in [15], MLP out-performs K-Nearest Neighbor (KNN) and SVM in Facebook trust prediction with accuracies of 83%, 73% and 71% respectively. In [16], Multinomial NB performs better than SVM. In [17], KNN, Naive Bayes, logistic regression and SVM are evaluated for performance. Each algorithm performed differently based on dataset and parameter selected. KNN and SVM were identified as having performed best with highest accuracies. In [18], the MLP performs better than Naïve Bayes at 93% to 88%.

It is common wisdom that there is no one predictive algorithm that is better than any other as performance of one depends on type/size of data and parameters selected. However, from literature review, SVM, Naïve Bayes and MLP probably stand out. In view of the fact that the SVM architecture used in most of the already-reviewed cases is optimized even as the MLP architecture used is un-optimized, our paper describes an attempt to explore whether an optimized MLP can perform better than the techniques reviewed here.

An MLP can have several hidden layers. Formally, an MLP having a single hidden layer forms the function:

$$f: R^D \rightarrow R^L,$$

where  $D$  is the size of input vector  $x$  and  $L$  is the size of the output vector  $f(x)$ , so that, in matrix notation:

$$f(x) = G(b^{(2)} + W^{(2)}(s(b^{(1)} + W^{(1)}x))) \quad (1)$$

with bias vectors  $b^{(1)}, b^{(2)}$ ; weight matrices  $W^{(1)}, W^{(2)}$  and activation functions  $G$  and  $s$ .

The vector

$$h(x) = \Phi(x) = s(b^{(1)} + W^{(1)}x) \quad (2)$$

forms the hidden layer.  $W^{(1)} \in R^{D \times D_h}$  is the weight matrix connecting the input vector to the hidden layer. Each column  $W_{\cdot i}^{(1)}$  represents the weights from the input units to the  $i^{\text{th}}$  hidden unit. Usual choices for  $s$  include *tanh*, with

$$\tan h(a) = (e^a - e^{-a}) / (e^a + e^{-a}) \quad (3)$$

, or the logistic *sigmoid* function, with

$$\text{Sigmoid}(a) = 1 / (1 + e^{-a}) \quad (4)$$

Both the *tanh* and *sigmoid* are scalar-to-scalar functions, however, their natural conversion to vectors and tensors is in applying them element-wise.

The output vector is therefore:

$$O(x) = G(b^{(2)} + W^{(2)}h(x)) \quad (5)$$

To train an MLP, all parameters of the model is learned. Using Stochastic Gradient Descent with mini-batches, the parameters to learn is in the set:

$$\Theta = \{W^{(2)}, b^{(2)}, W^{(1)}, b^{(1)}\} \quad (6)$$

Gradients  $\partial \ell / \partial \Theta$  is obtained using the back-propagation algorithm [19]. In our case, D shall be 2 and L shall be 1.

In order to identify and use the most optimal MLP model, several un-optimized or manually-trained MLPs can be tried, randomly varying the attributes of MLP such as number of neurons, number of layers, activation functions and learning rates. The best-performing of the various tested models is then used to conduct prediction. Optionally, automated technique can be used with one of two main optimizers; genetic algorithm (GA) and particle swarm optimization (PSO). The optimizer would automatically try out various values of the mentioned attributes out of a large solution space, each time testing the performance for most optimal model.

In GA, the problem would first be encoded as a string of real numbers or, as is more typically the case, a binary bit string. A typical chromosome may look like this:

1001010111010100101001110110

The following set of steps is a typical algorithm which is repeatable until an optimum solution is found.

- Test each chromosome to see how good it is at solving the problem at hand and assign a fitness score accordingly. The fitness score is a measure of how good that chromosome is at solving the problem to hand.
- Select two members from the current population. The chance of being selected is proportional to the chromosomes fitness. Roulette wheel selection is a commonly used method.
- Dependent on the crossover rate crossover the bits from each chosen chromosome at a randomly chosen point.
- Step through the chosen chromosomes bits and flip dependent on the mutation rate.
- Repeat step 2, 3, 4 until a new population of N members has been created.

For instance, given two chromosomes, one represented by black digits with the second one represented by grey digits;

10001001110010010  
01010001001000011

We can choose a random bit along the length, say at position 9, and swap all the bits after that point so the initial chromosomes above become:

10001001101000011  
01010001010010010

This would sometimes be followed by mutation. This is where a bit within a chromosome will be flipped (0 becomes 1, 1 becomes 0). This is usually a very low value for binary encoded genes, say 0.001. So whenever chromosomes are chosen from the population the algorithm first checks to see if crossover should be applied and then the algorithm iterates down the length of each chromosome mutating the bits if applicable[20] [21] [22]. In our case, the chromosomes of the GA would represent the number of neurons, the number of layers, the activation function and the learning rate.

PSO, on the other hand, is initialized with a group of random particles (solution) and then searches for optima in a conceptual 3D space by updating generations. In every iteration, each particle is updated by following two best 'values'. The first one is the best solution (fitness) it has achieved so far. The fitness value, called *pbest* is also stored. Another 'best' value that is tracked by the particle swarm optimizer, obtained so far by any particle in the swarm. This best value is a global best and called *gbest*. When a particle takes part of the population as its topological neighbors, the best value is a local best and called *lbest*. After finding the two best values, the particle updates its velocity and position with the following equations (1) and (2), respectively;

$$[v]=[v]+c1*rand()*(pbest[]-present[]) + c2*rand()*(gbest[]- present[]) \quad (1)$$

$$present[] = present[] + v[] \quad (2)$$

where  $v[]$  is the particle velocity;  $present[]$  is the particle or solution;  $pbest[]$  and  $gbest[]$  are defined as already-mentioned;  $rand[]$  is a random number between (0,1);  $c1$  and  $c2$  are learning factors, typically both equal to 2 [23].

Fig. 1 represents the operation of the MLP being optimized by either GA or PSO.

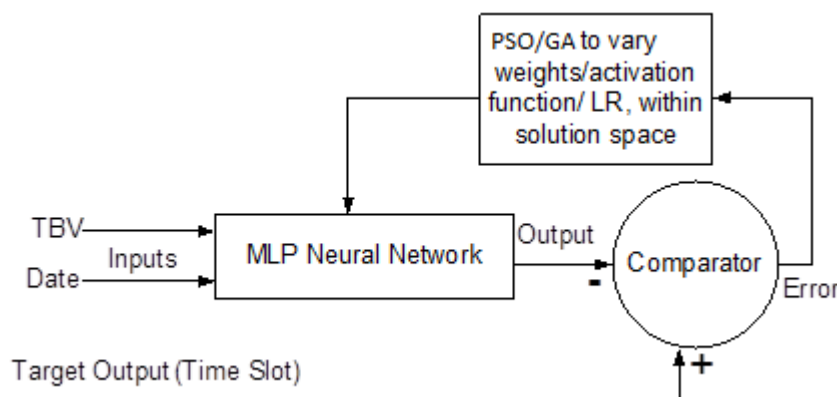


Figure 1 Logical Diagram for Process of Optimizing MLP

### III. METHODS

In order to model and predict the jam times for mobile telephony, we selected a perennially-jammed mobile cell for the dominant service provider, Safaricom Ltd. This is the kind of cell which mobile traffic jam times, one would



most desire to predict, as it would most benefit from opportunistic use of idle TV channels. From the transactional summary data recorded by the provider’s system within Nairobi city, we were able to identify Moi Drive cell as one of the worst cells in terms of traffic level and blocked calls rate. As earlier stated by Mahalungkar and also Verma, the main cause of blocked calls is shortage of frequency to allocate during jam times. Safaricom Ltd forensic data computed by the cell base station controller (BSC) also confirm that the call blocking rates are high due to shortage of frequency. Moi Drive cell is a logical choice although the central business district is the busiest area within the city with the highest mobile traffic. However, it has many tall buildings to anchor more towers thus enabling reduction of cell size. That measure is usually enough mitigation for high traffic. Moi Drive cell on the other hand, is a heavily populated, lower-middle-class and low-class, residential area with many busy open markets, but hardly any tall building. The provider has to literally build towers from the ground which is a costly affair. The cell, among others like it, suffers mobile traffic jam especially in late afternoon at the height of market activity. We obtained data transacted by the Base Station Controller (BSC) for the cell of study, data which is copied to the core system of the service provider. TABLE I, below illustrates the most relevant view of the transaction data;

TABLE I  
TRANSACTION DATA OF THE BSC

DATE	TIME	CI	SITEID	CELLNAME	BSC_ NAME	ACTUAL_ TRAFFIC	TRUE_ Blocking Value	Utilization Value	HR_ Proportion Value
2013-04-17	16.33	479	479-Umoja_Moi_Drive-6	BSC-1114	47.11	48.86	72.63	85.92	4796
.	.	.	.	.	.	.	.	.	.

As previously stated, to predict mobile traffic jam times, there is need to predict the times when there is highest call blocking rates. Of the fields in TABLE I, the relevant fields to model traffic jam are, therefore, True Blocking Value (TBV), which is Safaricom Ltd system’s term for traffic channel (TCH) blocking rate [24] [25]. Rate of TCH unsuccessful seizures during assignment procedure due to congestion is computed by the BSC as follows:

$$TCH\_Assignment\_Block\_Rate = \frac{TCH\_Blocks}{TCH\_Normal\_Attempts} \times 100\%$$

Of the fields in the data view, True Blocking Value, Day/Date and Time of Day are best suited to build the needed predictive model. The format of data for the three variables is illustrated in TABLE II, below;

TABLE II  
DATA FIELDS FOR MODELLING MOBILE TRAFFIC

DATE	TIME	TRUE_Blocking_Value
2013-04-17	16.33	72.63
.	.	.

There appears to be a weekly cyclic pattern in the data so that TBV spikes on Friday afternoons and relatively lowers on Monday afternoon than any other day of the week. We therefore converted the Date into Day-of-the-week. This conversion also spares us using 30 different symbols for the 30 days of the monthly date. We deal with 7 symbols for days of the week. The resulting data is therefore as follows;

TABLE III  
REVISED DATA FIELDS FOR MODELLING MOBILE TRAFFIC

DATE	TIME	TRUE_Blocking_Value
Wednesday	16.33	72.63
.	.	.

Variously-optimized multi-layer perceptron (MLP) models explored in this study work best with numeral data. Although it is tempting to give Monday a value of 1, Tuesday a value of 2, Wednesday a value of 3 etc., such data representation would not capture the implication of data as giving Friday a value of 5 may imply that Friday is 5 times the value of Monday which is not accurate. We used 1-of-N technique to encode data [26]. The resulting data is illustrated in Table IV, below;

TABLE IV  
ENCODED DATA FIELDS FOR MODELLING MOBILE TRAFFIC

DATE	TIME	TRUE_Blocking_Value
0,0,1,0,0,0,0	16.33	72.63
.	.	.

The BSC computes the data every 30 seconds. In order to avoid collecting too much data, we isolated and collected data of every 3 minutes. It was resolved that data of 30 days, which is 4 weeks, has enough cyclic patterns to enable prediction. The 14,400 instances were applied to different predictive algorithms. Given that in literature review, the MLP had rallied competitively against most sophisticated algorithms, we compared the MLP against its two optimized versions; Genetic Algorithm (GA)-optimized MLP and Particle Swarm Optimization (PSO)-optimized MLP. Version R2010a of MATLAB was used to develop the script to implement the MLP.

TABLE V illustrates the performance of the various models of a manually-trained MLP. Fig. 2 illustrates the training process and performance of the most optimal of the manually-trained MLP models. Fig. 3 illustrates the training process and performance of the GA-optimized MLP while Fig. 4 illustrates the training process and performance of the PSO-optimized MLP.

#### IV. RESULTS

TABLE V  
FOUR BEST MANUALLY-TRAINED MODELS OF MLP

No. of Hidden Layers	No. of Neurons	Learning Rate	Training Function	Mean Square Error
2	40	0.04	log sigmoid	0.198771
4	20	0.05	log sigmoid	0.026668
1	30	0.02	tan sigmoid	0.022933
3	20	0.4	tan sigmoid	0.083454

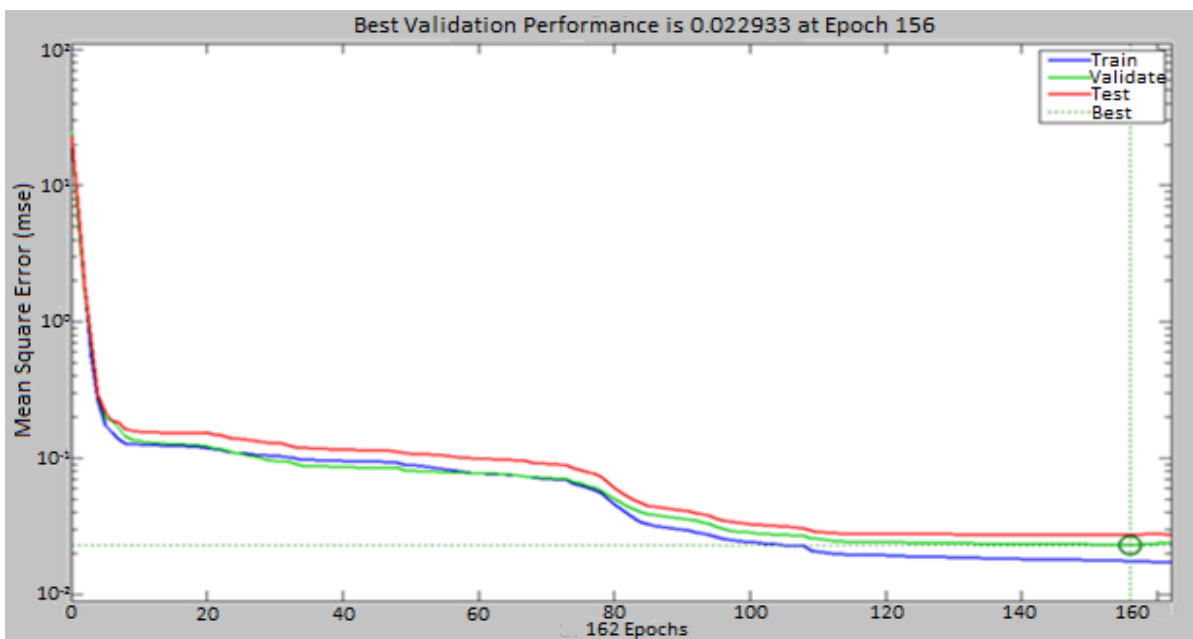


Figure 2 Training Graph of the best-performing MLP

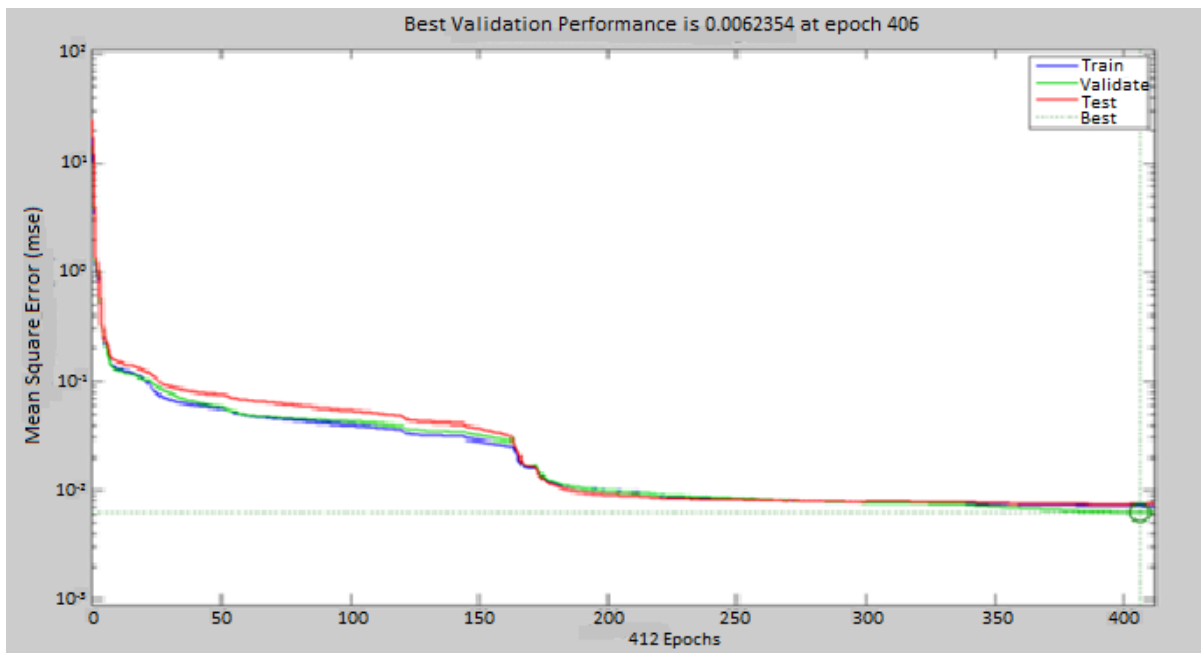


Figure 3 Training Graph of GA-Optimized MLP

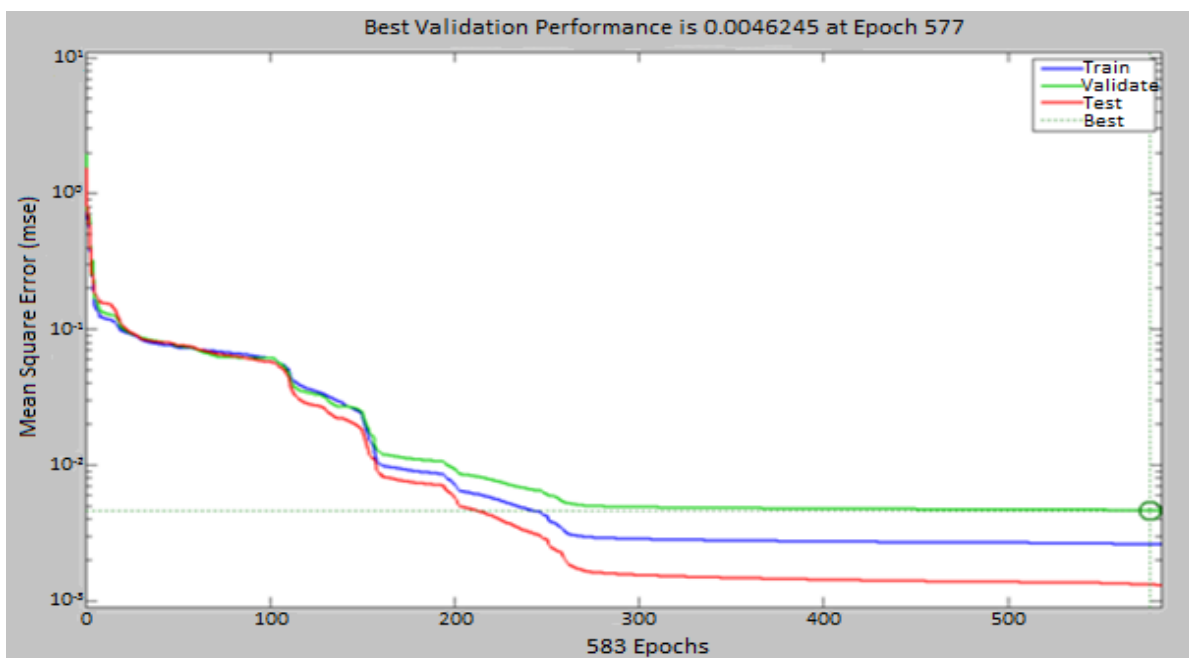


Figure 4 Training Graph of PSO-Optimized MLP

## V. DISCUSSION

When idle TV channels are used opportunistically by overwhelmed mobile telephony network, the benefit of predicting the times of mobile traffic jam is the ability to plan in advance on when to determine if idle frequency is available. In the attempt to predict the mobile telephony traffic jam times, we first compared the performance of several predictive algorithms by reviewing existing literature.

In one literature, wavelet transformation least squares support vector machines (LSSVM) is used to conduct busy telephone traffic prediction with impressive mean relative error of -0.0102. In another literature, a combination of ARIMA model and PSO-LSSVM is used to predict busy mobile traffic with a least error value of 1.14%. When in one study, Naïve Bayes, Bayesian Network and C4.5 decision tree were used to predict churn in mobile telephony. Naïve Bayes and Bayesian Network perform better than C4.5 decision tree with Naïve Bayes performing best with a relative error of -0.0456. In another case, the SVM of the Gaussian Radial basis Polynomial Function achieved an accuracy of 95.79% against 84.50% for the MLP. In one more case, SVM of the radial basis function achieved an accuracy of 100% against the MLP with 95%. In yet another case, the SVM performed better than MLP by an average of 19% over the 2008 to 2012 period. Still in another case, Multinomial NB performs better than SVM. In another case, KNN and SVM performed better than Naive Bayes and logistic regression.

Although in all the above circumstances, the SVM is seen to out-perform the MLP, the SVM is actually optimized while the MLP is non-optimized. Still, the MLP rallies well given that it still out-performs K-Nearest Neighbor (KNN) and SVM in Facebook trust prediction with accuracies of 83%, 73% and 71% respectively in one case. In yet another case, the MLP performs better than Naïve Bayes at 93% to 88%. Given that the MLP performs well without optimization, our study attempted to optimize the MLP using two popularly-used optimization algorithms; genetic algorithm (GA) and particle swarm optimization (PSO) and compare their performance with the un-optimized MLP and the other previously-reviewed algorithms.

In our study, the most optimal un-optimized or manually-trained MLP achieved an MSE of 0.022933 at epoch 156 with 1 hidden layer, 30 neurons, learning rate of 0.02 and training function of tan sigmoid. The GA-optimized MLP achieved even better with an impressive MSE of 0.0062354 at epoch 406. Still, the PSO-optimized MLP achieved the best performance with an MSE of 0.0046245 at epoch 577. Compared to the performances previously-reviewed, the PSO-optimized MLP has so far the best performance.

## VI. CONCLUSION

The study can conclude that mobile telephony traffic in one of Safaricom Ltd's perennially-jammed mobile cell in Nairobi city has predictable patterns.

The study can also conclude that the manually-trained MLP has competent performance against other traditionally top-performing predictive algorithms such as SVM, Naïve Bayes, K-Nearest Neighbor and Decision trees.

The study can finally conclude that with this particular data set, the PSO-optimized MLP probably would perform better than all the previously-reviewed algorithms.

## REFERENCES

- [1] S. Pociask, JUN 30, 2015. "We're Three Year Away From Spectrum Shortages". Forbes. <http://www.forbes.com/sites/realspin/2015/06/30/the-spectrum-shortage-is-coming/>
- [2] Z. Feng, 2014. Cognitive Cellular Systems in China Challenges, Solutions and Testbed. ITU-R SG 1/WP 1B WORKSHOP: SPECTRUM MANAGEMENT ISSUES ON THE USE OF WHITE SPACES BY COGNITIVE RADIO SYSTEMS (Geneva, 20 January 2014). <http://www.itu.int/en/ITU-R/study-groups/workshops/RWP1B-SMWSCRS-14/Presentations/CHN%20-%20Cognitive%20Cellular%20Systems%20in%20China.pdf>
- [3] G. P. Joshi, Seung Yeob Nam, and Sung Won Kim, 2013. Cognitive Radio Wireless Sensor Networks: Applications, Challenges and Research Trends. Sensors (Basel). 2013 Sep; 13(9): 11196–11228. Published online 2013 Aug 22. doi: 10.3390/s130911196
- [4] Communications Authority of Kenya (2014) Annual Report for the Financial Year 2012-2013. [Online] Pp 36. Available: <http://www.ca.go.ke/index.php/annual-reports> [Jul. 2014]
- [5] S. P. Mahalungkar, and S. S. Sambare, (2012) Call Due to Congestion in Mobile Network. Journal of Computer Applications (JCA) ISSN: 0974-1925, Volume V, Issue 1, 2012
- [6] P. Verma, P. Sharma, and S. K. Mishra, (2012) Dropping of Call Due to Congestion in Mobile Network. Journal of Computer Applications (JCA) ISSN: 0974-1925, Volume V, Issue 1, 2012
- [7] ITU (2012) Agenda and References; Resolutions and Recommendations. World Radio Communication Conference 2012 (WRC-12)
- [8] B. R. Danda, S. Min, and S. Shetty, 04 February 2015. Resource Allocation in Spectrum Overlay Cognitive Radio Networks. Book - Dynamic Spectrum Access for Wireless Networks. Part of the series SpringerBriefs in Electrical and Computer Engineering pp 25-42. 10.1007/978-3-319-15299-8\_3
- [9] J. Li, Z. Jia, X. Qin, L. Sheng, and L. Chen, 2013. Telephone Traffic Prediction Based on Modified Forecasting Model. Research Journal of Applied Sciences, Engineering and Technology 6(17): 3156-3160, 2013 ISSN: 2040-7459; e-ISSN: 2040-7467 © Maxwell Scientific Organization, 2013
- [10] W. He, X. Qin, Z. Jia, C. Chang, and C. Cao, 2014. Forecasting of Busy Telephone Traffic Based on Wavelet Transform and ARIMA-LSSVM. International Journal of Smart Home Vol.8, No.4 (2014), pp.113-122 <http://dx.doi.org/10.14257/ijsh.2014.8.4.11>
- [11] C. Kirui, L. Hong, W. Cheruiyot, and H. Kirui, 2013. Predicting Customer Churn in Mobile Telephony Industry Using Probabilistic Classifiers in Data Mining. IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 2, No 1, March 2013 ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784 [www.IJCSI.org](http://www.IJCSI.org)
- [12] E. A. Zanaty, 2012. Support Vector Machines (SVMs) versus Multilayer Perception (MLP) in data classification. Egyptian Informatics Journal. Volume 13, Issue 3, November 2012, Pages 177–183. doi:10.1016/j.eij.2012.08.002
- [13] M. C. Lee, and C. To, 2010. Comparison of Support Vector Machine and Back Propagation Neural Network in Evaluating the Enterprise Financial Distress. International Journal of Artificial Intelligence & Applications (IJAIA), Vol.1, No.3, July 2010. DOI: 10.5121/ijaia.2010.1303
- [14] J. K. Mantri, 2013. Comparison between SVM and MLP in Predicting Stock Index Trends. International Journal of Science and Modern Engineering (IJSME) ISSN: 2319-6386, Volume-1, Issue-9, August 2013
- [15] E. Khadangi, and A. Bagheri, 2013. Comparing MLP, SVM and KNN for predicting trust between users in Facebook. Computer and Knowledge Engineering (ICCKE), 2013 3th International eConference. Oct. 31 2013-Nov. 1 2013. Page(s): 466 – 470. Conference Location: Mashhad. DOI: 10.1109/ICCKE.2013.6682864
- [16] S. Matwin, and V. Sazonova, 2012. Direct comparison between support vector machine and multinomial naive Bayes algorithms for medical abstract classification. Journal of the American Medical Informatics Association. J Am Med Inform Assoc. 2012 Sep-Oct; 19(5): 917. doi: 10.1136/amiajnl-2012-001072
- [17] M. Rana, P. Chandorkar, A. Dsouza, and N. Kazi, 2015. BREAST CANCER DIAGNOSIS AND RECURRENCE PREDICTION USING MACHINE LEARNING TECHNIQUES. IJRET: International Journal of Research in Engineering and Technology eISSN: 2319-1163 | pISSN: 2321-7308
- [18] S. A. Kumar, P. S. Kumar, and A. Mohammed, 2014. A Comparative Study between Naïve Bayes and Neural Network (MLP) Classifier for Spam Email Detection. International Journal of Computer Applications® (IJCA) (0975 – 8887) National Seminar on Recent Advances in Wireless Networks and Communications, NWNC-2014
- [19] S. Haykin (1998). *Neural Networks: A Comprehensive Foundation (2 ed.)*. Prentice Hall. ISBN 0-13-273350-1. (1998). pp. 34-57
- [20] J. Jiang (2013). "BP Neural Network Algorithm Optimized by Genetic Algorithm and its Simulation." *International Journal of Computer Science* [Online] Issues vol. 10, Issue 1. No. 2. ISSN: 1694-0814 [Oct. 2014]
- [21] G. Panchal, and A. Ganatra, (2012). *Optimization of Neural Network Parameter Using Genetic Algorithm: Extraction of Neural Network Weights Using GA-based Back Propagation Network (2<sup>nd</sup> Ed)*. LAP LAMBERT Academic Publishing. ISBN-13: 978-3848447473. (2012) pp. 123 and 136
- [22] G. Heath. (2013). "GA Optimization of NN Weights." Internet: [http://www.mathworks.com/matlabcentral/newsreader/view\\_thread/326543](http://www.mathworks.com/matlabcentral/newsreader/view_thread/326543). 2013 [Oct. 2013]
- [23] A. Espinal, M. Sotelo-Figueroa, J.A. Soria-Alcaraz, M. Ornelas, H. Puga, M. Carpio, R. Baltazar, and J.L Rico, 2011. Comparison of PSO and DE for Training Neural Networks Artificial Intelligence (MICAI), 2011 10th Mexican International Conference on Page(s): 83 – 87
- [24] L. Harte, B. Bramley, and M. Davis, (2012) Introduction to GSM: Physical Channels, Logical Channels, Network Functions, and Operation, 3<sup>rd</sup> Edition. Pp 31-85. Althos Publishing
- [25] J. Eberspächer, H. Vögel, C. Bettstetter, and C. Hartmann, (2009). GSM - Architecture, Protocols and Services Hardcover. 3<sup>rd</sup> Edition. ISBN-13: 978-0470030707. ISBN-10: 0470030704
- [26] W. D. Mulder, S. Bethard, and M. F. Moens, 2014. A survey on the application of recurrent neural networks to statistical language modeling. Computer Speech & Language Volume 30, Issue 1, March 2015, Pages 61–98 ELSEVIER. doi:10.1016/j.csl.2014.09.005





**First W. Ojenge** was born in 1969 in Kisumu District, Kenya. His M.Sc. was in information systems (artificial intelligence) from the University of Nairobi in 2008. He is pursuing the Ph.D. in Computer Science at Technical University of Kenya, Nairobi.

From 2009 to present, he has taught electrical engineering and computer science at the Technical University of Kenya and Strathmore University. His research interests include machine learning application in telecommunications, robotics and Internet of Things.

Mr. Ojenge has authored toward three IEEE conferences in the last two years. The papers appear in the IEEE Xplore Digital Library.



**Second W. Okelo-Odongo** was born in Kisumu District, Kenya, in 1953. Between 1975 and present, he has received: the B.Sc. in Mathematics/Computer Science with highest honors from Northwest Missouri State University, Missouri, USA; M.Sc. in Electrical Engineering with concentration in Computer and Communication Systems from Stanford University, California; Ph.D. in Computer Science from the University of Essex, U.K; been a Computer programmer with Roberts and Dybdahl, Inc., Iowa, USA; a Project engineer with EG&G Geometrics, Inc., Sunnyvale California, USA; part of faculty in the School of Computing and Informatics, University of Nairobi, becoming an Associate Professor in 2006; the Director, School of Computing and Informatics, University of Nairobi. Within the school, between 2009 and 2013,

he has been the Project coordinator for UNESCO-HP Brain Gain project and ITES/BPO Project. He has been teaching at undergraduate and postgraduate levels for over 20 years, and has supervised many M.Sc. students plus 3 Ph.D. students to completion. He is currently supervising 7 ongoing Ph.D. students. He has authored over 30 publications and his research areas of interest are distributed computing including application of mobile technology, computing systems security and real-time systems.

Prof. Okelo-Odongo is a member of the Kenya National Council for Science and Technology Physical Sciences Specialist Committee and ICT Adviser to the Steering Committee of *AfriAfya* Network: An ICT for community health project sponsored by the Rockefeller foundation. He is also a member of The Internet Society (ISOC).



**Third P.J. Ogao** was born in 1967 in Tabora, Tanzania. Between 1990 and present, he has: obtained a degree in Surveying and Photogrammetry from the university of Nairobi; obtained an MSc in Integrated Map and Geo-information Production from the International Institute for Geo-information Science and Earth Observations in Enschede; obtained the PhD in Geo-informatics from Utrecht University, Netherlands; lectured in several universities, including: University of Groningen, The Netherlands; Kyambogo, Mbarara and Makerere Universities in Uganda; Masinde Muliro University and Technical University of Kenya. He has gathered 12 professional development certificates from the USA, UK, France and The Netherlands; supervised several PhD students; done numerous publications including a book; *Exploratory Visualization of Temporal Geospatial Data Using Animation*, ISBN 90-6164-206-X. His research interests are in visualization applications in bio-informatics, geo-informatics, and software engineering and in developing strategies for developing countries.

Prof. Ogao is a recipient of the following awards and scholarships: ICA Young Student Award, Ottawa, Canada; European Science Foundation Scholar, Ulster, UK; ESRI GIScience Scholar, USA; ITC/DGIS, PhD Fellowship; Netherlands Fellowship Programme, MSc Research Fellowship, The Netherlands; Survey of Kenya-IGN-FI Training Award, Paris, France. He is also a member of ACM SIGGRAPH; Associate member, Institution of Surveyors, Kenya; Member of Commission for Visualization and Virtual Environment; Member of Research Group of Visualization and Computer Graphics, University of Groningen, The Netherlands.

# New Variant of Public Key Based on Diffie-Hellman with Magic Cube of Six-Dimensions

Ph. D Research Scholar Omar A. Dawood<sup>1</sup>

Prof. Dr. Abdul Monem S. Rahma<sup>2</sup>

Asst. Prof. Dr. Abdul Mohsen J. Abdul Hossen<sup>3</sup>

Computer Science Department

University of Technology, Baghdad, Iraq

**Abstract-** In the present paper we are developed a new variant of asymmetric cipher (Public Key) algorithm that based on the Diffie-Hellman key exchange protocol and the mathematical foundations of the magic square and magic cube as the alternative to the traditional discrete logarithm and the integer factorization mathematical problems. The proposed model uses the Diffie-Hellman algorithm just to determine the dimension of magic cube's construction and through which determines the type of based magic square in the construction process if it is (odd, singly-even or doubly-even) type, as well as through which determined the starting number and the difference value in addition to the face or dimension number that will generate the ciphering key to both exchanged parties. From the other point it exploits the magic cube characteristics in encryption/decryption and signing/verifying operations. The magic cube is based on the folding six of series magic squares with sequential or with period numbers of n-dimensions that represent the faces or dimensions of magic cube. The proposed method speed up the ciphering and deciphering process as well as increases the computational complexity and gives a good insight to the designing process. The magic sum and magic constant of the magic cube play a vital role in encryption and decryption operations that imposes to keep as a secret key.

**Keywords:** Magic Cube, Magic Square, Diffie-Hellman, RSA, Digital Signature.

## I. INTRODUCTION

Magic squares remain an interesting phenomenon to be studied, both mathematically and historically. It is equivalent to a square matrix as a painting full of numbers or letters in certain arrangements. Mathematics is the most interesting subject in computational squares consisting of  $n^2$  boxes, called cells or boxes, filled with different integers [1]. This array is called magic square of  $n \times n$  numbers containing the numbers with consecutive order as  $1; 2 \dots n^2$ . The total elements in any row, column, or diagonals should be the same [2]. Therefore; Magic cube is an extension to the magic square with three dimensions or more, that contains an arrangement set of integer number from  $1, 2, \dots, n^3$ . The sum of the entries elements in rows direction, columns direction, and all the diagonals direction give the same magic constant for the cube. A magic cube construction of order 3 is shown in Figure 1. below [3].

10	26	6
24	1	17
8	15	19

23	3	16
7	14	21
12	25	5

9	13	20
11	27	4
22	2	18

Figure 1. Magic Cube of Order Three

The magic cube is like the magic square from the point of probability construction that increases dramatically with the order of magic cube but with higher search space in guessing and estimation. The magic cube in Figure 2. below is another direction in the cube construction. The starting element in the diagonal cube begins from one corner of the cube that comprises the upper layer dimensions through the lower left corner. This is the smallest normal magic cube of 3x3x3 dimensions with sequential numbers from 1 to 27 that are organized in three layers of nine numbers and the magic constant for this cube is sum to 42. These layers represent the dimension or face for the magic cube that arranged magically from all directions [4].

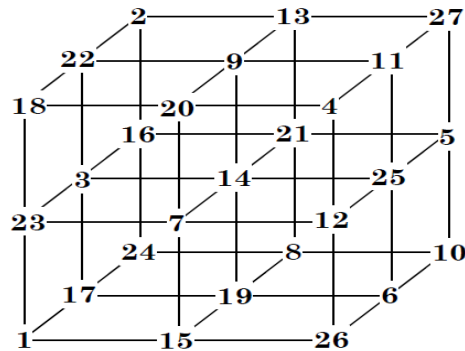


Figure 2. Magic Cube of Three Layers

Magic cubes are more than playing games with numbers like the chessboard or Rubik cube, but they substantially depend upon the mathematical rules in their construction. Magic cubes are embedded in several mathematical fields like the number theory, matrices and combinatory etc. [5].

## II. RELATED WORKS

The embedding of magic square & magic cube techniques through the public key cipher still under development and research and it is still evolving, since there is no real encryption and decryption method with full dependence on magic square principles and comprehensions, because it uses the magic square just as the additional security layer and as an assistant tool for the real asymmetric algorithms.

In [6] **Nitin Pandey and D.B.Ojha** have proposed a new method of encryption and decryption process that is based on the RSA cipher and magic rectangle's construction. The proposed method constructs different of singly even magic of rectangles with an even order that don't accept the divisibility by four, where the sum of each rows and columns values is the same. The main purpose for the rectangle square is to address the numeral values to the corresponding positions with the magical rectangle in different quadrants. So the numbers then encrypted and decrypted using the RSA public cipher. The two researchers have proposed that the developed method increases the complexity and the randomness for the ciphertext and at the same time it requires an extra time for the implementation process.

In [7] **A. Dharini, R.M. Saranya Devi, and I. Chandrasekar** have introduced a new approach for secure data transmission through the cloud environment and sharing networks as well as during the Secure Socket Layer (SSL) by the RSA combined with magic square, to provide additional security layer to the cryptosystem. The proposed

model submits the confidentiality and the integrity of data over the communication to and from the cloud providers. So it discusses and combines magic square algorithm with the RSA cipher when implemented on data security in cloud computing.

**Gopinath, Ganapathy, and K. Mani** have proposed a new approach for the public key cipher with magic square which is based on generating the magical square with order of doubly even for multiples of sixteen. The developed method is not much different from the previous work except it has taken various methods of the public key cipher. Therefore; the proposed method treated with the positions of the magic square that are corresponding to the ASCII values as an alternative. So the encryption and decryption will be with different numbers that represent the placements of the ASCII elements in the magic square. The ciphering and deciphering process is performed also by the RSA algorithm [8].

**D.I. George, J.Sai Geetha and K.Mani** are three researchers from India who have proposed another technique of combining the RSA cipher with the magic rectangle of singly even order. The construction of magical rectangle is based on several initial parameters that involve magical rectangle's seed, constant vector for the amount of column values and the beginning number for the rectangle construction. The main idea for this method is to construct a specific magic rectangle and then the encrypted text or the cipher text which is encrypted by the RSA cipher and then mapping to the positions of numeric values in the magic rectangle and after that change the ciphertext with those corresponding positions in order to increase the time complexity and to add another security level. The proposed method requires more time of implementation and hardware cost [9].

**In 1970, Richard Meyers** has invented a perfect eighth-order magic cube that is known by Meyers cube. The Meyers cube is interested in several symmetries properties which assume that the cube is associative and every orthogonal and diagonal line sum to the same specific number. The corners values in the inner small cube as well as the corners values of each rectangular in the Meyers cube also sum to a constant certain number. The prominent feature for the symmetries properties makes that is possible for a tantalizing number and for rearrangements of the cube [10].

**In 1981 J. Barkley Rosser and Robert J. Walker** are two researchers who have introduced a new approach for constructing a perfect eighth-order magic cube. They have also explained and proved that the perfect pan diagonal cubes are found for whole orders with multiples of 8 and also for all the odd orders that are more than 8 order [11].

**In 1988, John Hendricks** submitted new ideas and published many refereed related papers. He developed a simple and clear technique in constructing of an odd order magic cube with N order. In addition, he published an extended dimension of the hyper cubes with four, five and six dimensions; also he applied an elegant work and great share in the magic square area and in the methods development for the magical constructions approaches [12].

### III. DIFFIE-HELLMAN KEY EXCHANGE

The Diffie-Hellman is one of the first and earliest public-key protocols that allows two parties to share a secret key without any predated acquaintance over insecure or untrusted channel. The resultant key can be used later to encrypt as a symmetric key cipher. The Diffie-Hellman protocol is based mainly on the Discrete Logarithm Problem (DLP) and implemented within the finite field of cryptography. Suppose Alice and Bob want to share a secret key over

public communication media and to use then in a symmetric cipher, so the information and the key that they exchange will be observed by their adversary Eve [13]. At the beginning Alice and Bob have to agree on a finite field ( $F_q$ ) and the based element ( $g$ ) in  $F_q$ , then each one secretly chooses a random positive number  $n$  and  $m$  to compute the following:

$$\text{Alice computes } X \equiv g^n \pmod{p}$$

$$\text{Bob computes } Y \equiv g^m \pmod{p}$$

The two parties will calculate the share key, where Bob raises the base element to the selected value that has been received from Alice, also the Alice raises the base element to the value that has been received from Bob to complete the agreement on the shared secret key as stated below:

$$\text{Alice computes } X \equiv Y^n \pmod{p}$$

$$\text{Bob computes } Y \equiv X^m \pmod{p}$$

Then, the two parties shall have  $(g^n)^m = (g^m)^n = g^{nm} \in F_q^*$ .

The Diffie-Hellman protocol suffers from the man-in-the-middle attack which is considered a form of eavesdropping attack that happens when the malicious attacker or eavesdropper monitors, modifies and retransmits the intercepted information across the communication session between the two users by impersonating the personality of the authorized author. The Diffie-Hellman concepts paved the way to the invention the RSA public cipher [14].

#### IV. THE PROBABILITY OF THE MAGIC SQUARE AND MAGIC CUBE

The probability of construction the magical square increases considerably with the increasing in order of magic square as was mentioned earlier. So, there is merely one ordinary magic square of the third order, but by doing some reflections, transpositions and rotations, it will get seven other undifferentiated cases of magic square as stated in Figure 3. [15].

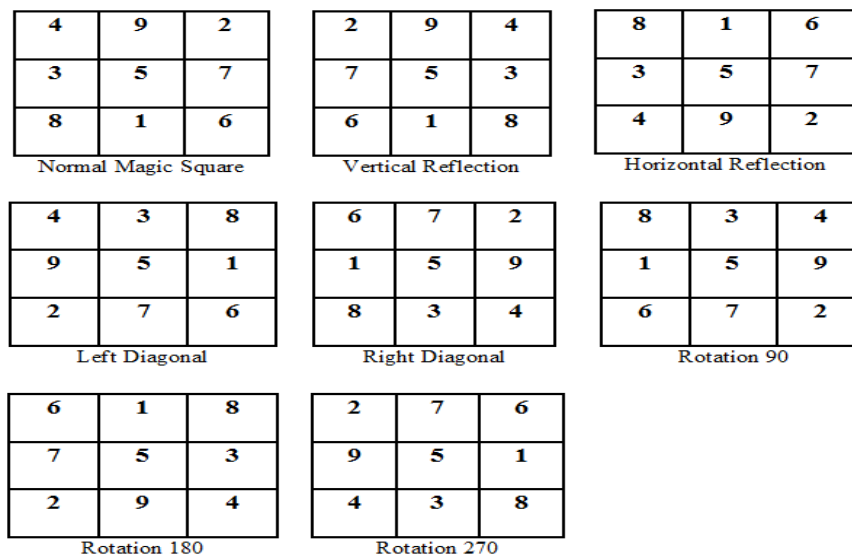


Figure 3. Rotations and Reflections of Magic Square

The probability of constructing magic square of fourth order (4x4) is 880 time, and the probability of constructing fifth order is more than 13 million normal magic squares. There are  $(n^2)!$  Styles to overfill the  $n \times n$  square matrix with integer's numbers between 1 and  $n^2$ , without redundancy, but there are just a few of them superpose the magic squares features. The construction of magic square acts a great challenge for the intelligent search methods for existing various magic square [16].

## V. THE CONSTRUCTION OF MAGIC CUBE

The construction of magic cube is the most difficult problem that has become interesting to researchers in mathematical sciences for a long time. Therefore; the methods that work for an odd order of magic cube will not work for doubly even or singly even methods and vice versa. The proposed method works for all types of magic cube and with any order and it depends basically on the magic square techniques. The proposed method allows to construct several magic cubes with sequential numbers or with constant differences among the series of numbers. The work for six squares (surfaces) will give one magic cube and the work for twelve squares will constitute two magic cubes and so on, so the work with cube should be multiplied of six numbers to introduce several cubes regarding the need or to the task requirements. The following example explains the core notation for the magic cube construction.

1. At the beginning, build six separated magic squares of any order corresponding to the six surfaces of the cube dimensions as shown in Figure 4. with order three.
2. Arrange the six surfaces (squares) of the cube by the following way: the first surface should be put opposite to the sixth surface and the second surface opposite to the fifth surface and finally places the third surface opposite to the forth surface with the corresponding colours respectively.

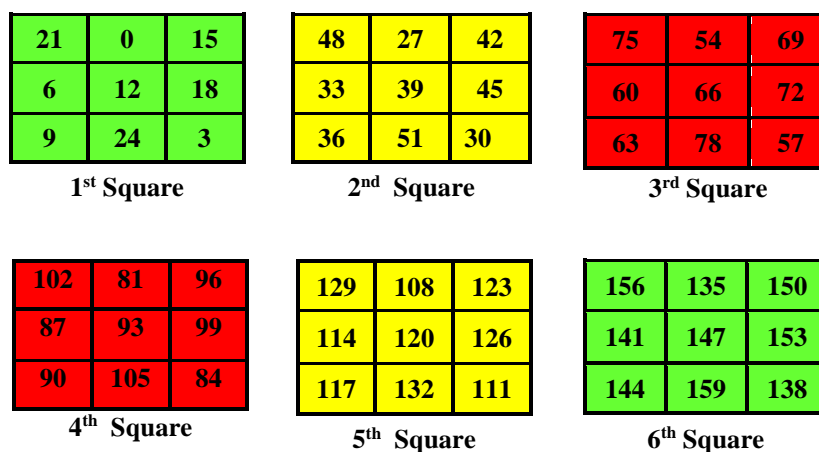


Figure 4. Six Faces of Folding Magic Cube

After constructing and coloring the six surfaces (dimensions) for the magic cube as it is illustrated in Figure 4. above, the magic constant and the magic sum are computed for each square consecutively. The summation for each pair of analogues colored square will give the same result.



## VI. THE PROPOSED MAGIC CUBE

The proposed method includes constructing a magic cube by using the folded magic square technique and it is considered a new step towards the magic cube construction that applies a good insight and provides an easy generalized technique. This method generalizes the design of magic cube with N order regardless of the type of magic square whether odd order or even order squares. The proposed method is fairly easy, since it depends mainly on the magic square construction methods, and all what the designer needs is just how to build six magic square sequentially or with constant difference value between each pair of the numbers in the square matrix, whereby each one of this magic square will represent the surface or dimension for the magic cube configuration. The next step for the designer will be how to arrange each square in the proper order to constitute the regular cube in order to maintain the properties of magic cube, where the sum of rows, columns and the diagonals from all directions are the same.

Magic cubes are more than playing games with numbers like the chessboard or Rubik cube, but they substantially depend upon the mathematical rules in their construction. Magic cubes are embedded in several mathematical fields like the number theory, matrices, and combinatorics. There exist eleven of distinct flat shapes that can be folded-up to construct a shape of cube as they were mentioned in the previous chapter. These shapes have been coloured with three distinctive colours (green, yellow and orange) each pair of opposite sides are coloured with same colour to constitute a folded cube with six surfaces each two opposite surfaces with the same colour as shown in Figure 5. below. The purpose of these colours is to keep the arrangement of the magic square as we shall explain in the following sections.

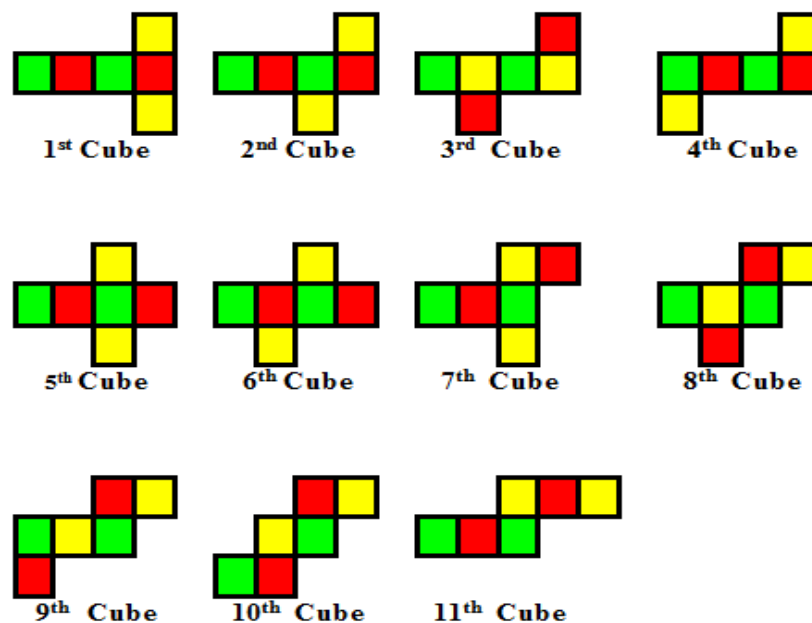


Figure 5. Different Shapes of Unfolded Cubes

## VII. THE CROSS FIGURE OF FOLDED CUBE

The cross shape that has been stated in Figure 6. below acts as one of the several shapes that construct the magic cube, selected to be traversed and tested with vertical and horizontal track for the main diagonals values and the

secondary diagonals values, from all directions which will produce the same constant number of (954). There is no problem or restricted matter in the selection of any shape from the magic cube's shapes else cross shape, where the cross shape is chosen because it is very easy, simple and clear to figure out the procedures of folding in the magic square surfaces. The cross shape will be partitioned into several basic parts in order to study its properties in detail.

			21	0	15			
			6	12	18			
			9	24	3			
48	27	42	75	54	69	129	108	123
33	39	45	60	66	72	114	120	126
36	51	30	63	78	57	117	132	111
			156	135	150			
			141	147	153			
			144	159	138			
			102	81	96			
			87	93	99			
			90	105	84			

Figure 6. Unfolded Magic Cube

The cubes' traversing and navigation checking will comprise that the first part will include the main and secondary diagonals for the matrices of magic cube with circular movement in both directions vertically and horizontally. The second part of the cube will involve the rows and columns of magic cube with circular movement in both directions vertically and horizontally and the traversing process that includes the tracks of values with rows and columns of the magic cube from all directions. These operations also will produce the same constant number of (954). The traversing process involves the tracks values with rows and columns of the magic cube from all directions. These operations will also produce the same constant number of (954).

#### VIII. APPLICATIONS OF THE PROPOSED MAGIC CUBE

Magic cube has no direct usage or specific applications. Recently, several published research papers embed the magic square and magic cube properties in many aspects and in several contributions such as:

1. Cryptography and Information Security.
2. Public Key and Secret Sharing.
3. Remote Access Control.
4. Applied Mathematics.
5. Number Theory.
6. Determinants and Matrices Field.
7. Coding theory and error Correctness
8. Game and the Search Algorithms

## IX. THE PROPOSED CIPHER

The proposed cipher involves that the two parties or more can use the proposed model to encrypt and decrypt the clear text using one of the famous symmetric algorithms cipher and then encrypts and sends the cipher key in secure form by asymmetric cipher using the proposed model that includes the combined of Diffie-Hellman algorithm with magic cube technique as well as signing and verifying the intended message by the same process by depends upon the Secure Hash Algorithm SHA-512 bits. The example below simulate the proposed cipher with step by step implementation. Figure 7. at the end of this paper exhibits the main flowchart for the whole operations of Exchanging, Encryption/Decryption and Signing/Verifying processes .

### Network Space

Diffie-Hellman Key Exchange ...  
Starting...      The Handshaking & Key Exchange

- (Alice and Bob Handshaking)

The Handshaking process is used just to ensure that the two parties are existing and available on the network and to prevent the man in the middle attack by depending upon one of the famous authentication protocols. The string below represents a simple simulation for the designed program for the handshaking session.

**Alice Sends:**

**39ZQKRYMOZ7C5G5LZDY7STQV9G3RB7JCS0RX6KBH1US4H0S2OH|5|31PFQKPILDN92TRLC3HY2  
P99V58KH9QZ98505LOZJXRO68903J**

**Bob Responds:**

**2DN194JXL3T4I50ICPUJDZF2UH1NRZIZOKGXH23D3ASXL7YKNC**

- (Alice and Bob Key Exchange)

The key exchange is implemented and generated by the Diffie-Hellman algorithm. The designed program treats with several hundreds of prime numbers that are generated and selected randomly to constitute the agreement key for the two parties module for a specific primitive element. The string below represents the generated key module the prime number of p. Where (P =1999).

**Alice key:**

**201069825420017342372071318209105168150744856972113247217941246143113212165129**

**Bob key:**

**201069825420017342372071318209105168150744856972113247217941246143113212165129**

The dimension of the constructed cube in this example will involve only the first byte from the generated key for easy calculation, since the designed program can take unlimited number of order n\*n. Thus, the dimension of cube constructing will be equal to 20.

Press any key...  
 Cube or square or print either c or s character...  
 c  
 Enter the dimension...  
 Enter the dimension is generated...20  
 Enter the lower range: 20  
 Enter the upper range: 222  
 Enter the period: 5  
 Enter the multiplied value: 3  
 Magic Cube with Period  
 Press any key to construct the cube with six dimensions consequently...

**The First Dimension**

425	27	28	422	421	31	32	418	417	35	36	414	413	39	40	410	409	43	44	406
46	404	403	49	50	400	399	53	54	396	395	57	58	392	391	61	62	388	387	65
66	384	383	69	70	380	379	73	74	376	375	77	78	372	371	81	82	368	367	85
365	87	88	362	361	91	92	358	357	95	96	354	353	99	100	350	349	103	104	346
345	107	108	342	341	111	112	338	337	115	116	334	333	119	120	330	329	123	124	326
126	324	323	129	130	320	319	133	134	316	315	137	138	312	311	141	142	308	307	145
146	304	303	149	150	300	299	153	154	296	295	157	158	292	291	161	162	288	287	165
285	167	168	282	281	171	172	278	277	175	176	274	273	179	180	270	269	183	184	266
265	187	188	262	261	191	192	258	257	195	196	254	253	199	200	250	249	203	204	246
206	244	243	209	210	240	239	213	214	236	235	217	218	232	231	221	222	228	227	225
226	224	223	229	230	220	219	233	234	216	215	237	238	212	211	241	242	208	207	245
205	247	248	202	201	251	252	198	197	255	256	194	193	259	260	190	189	263	264	186
185	267	268	182	181	271	272	178	177	275	276	174	173	279	280	170	169	283	284	166
286	164	163	289	290	160	159	293	294	156	155	297	298	152	151	301	302	148	147	305
306	144	143	309	310	140	139	313	314	136	135	317	318	132	131	321	322	128	127	325
125	327	328	122	121	331	332	118	117	335	336	114	113	339	340	110	109	343	344	106
105	347	348	102	101	351	352	98	97	355	356	94	93	359	360	90	89	363	364	86
366	84	83	369	370	80	79	373	374	76	75	377	378	72	71	381	382	68	67	385
386	64	63	389	390	60	59	393	394	56	55	397	398	52	51	401	402	48	47	405
45	407	408	42	41	411	412	38	37	415	416	34	33	419	420	30	29	423	424	26

Magic Constant =4510

Magic Sum =90200

**The Second Dimension**

430	32	33	427	426	36	37	423	422	40	41	419	418	44	45	415	414	48	49	411
51	409	408	54	55	405	404	58	59	401	400	62	63	397	396	66	67	393	392	70
71	389	388	74	75	385	384	78	79	381	380	82	83	377	376	86	87	373	372	90
370	92	93	367	366	96	97	363	362	100	101	359	358	104	105	355	354	108	109	351
350	112	113	347	346	116	117	343	342	120	121	339	338	124	125	335	334	128	129	331
131	329	328	134	135	325	324	138	139	321	320	142	143	317	316	146	147	313	312	150
151	309	308	154	155	305	304	158	159	301	300	162	163	297	296	166	167	293	292	170
290	172	173	287	286	176	177	283	282	180	181	279	278	184	185	275	274	188	189	271
270	192	193	267	266	196	197	263	262	200	201	259	258	204	205	255	254	208	209	251
211	249	248	214	215	245	244	218	219	241	240	222	223	237	236	226	227	233	232	230
231	229	228	234	235	225	224	238	239	221	220	242	243	217	216	246	247	213	212	250
210	252	253	207	206	256	257	203	202	260	261	199	198	264	265	195	194	268	269	191
190	272	273	187	186	276	277	183	182	280	281	179	178	284	285	175	174	288	289	171
291	169	168	294	295	165	164	298	299	161	160	302	303	157	156	306	307	153	152	310
311	149	148	314	315	145	144	318	319	141	140	322	323	137	136	326	327	133	132	330
130	332	333	127	126	336	337	123	122	340	341	119	118	344	345	115	114	348	349	111
110	352	353	107	106	356	357	103	102	360	361	99	98	364	365	95	94	368	369	91
371	89	88	374	375	85	84	378	379	81	80	382	383	77	76	386	387	73	72	390
391	69	68	394	395	65	64	398	399	61	60	402	403	57	56	406	407	53	52	410

50 412 413 47 46 416 417 43 42 420 421 39 38 424 425 35 34 428 429 31

Magic Constant =4610 Magic Sum =92200

The Third Dimension

435	37	38	432	431	41	42	428	427	45	46	424	423	49	50	420	419	53	54	416
56	414	413	59	60	410	409	63	64	406	405	67	68	402	401	71	72	398	397	75
76	394	393	79	80	390	389	83	84	386	385	87	88	382	381	91	92	378	377	95
375	97	98	372	371	101	102	368	367	105	106	364	363	109	110	360	359	113	114	356
355	117	118	352	351	121	122	348	347	125	126	344	343	129	130	340	339	133	134	336
136	334	333	139	140	330	329	143	144	326	325	147	148	322	321	151	152	318	317	155
156	314	313	159	160	310	309	163	164	306	305	167	168	302	301	171	172	298	297	175
295	177	178	292	291	181	182	288	287	185	186	284	283	189	190	280	279	193	194	276
275	197	198	272	271	201	202	268	267	205	206	264	263	209	210	260	259	213	214	256
216	254	253	219	220	250	249	223	224	246	245	227	228	242	241	231	232	238	237	235
236	234	233	239	240	230	229	243	244	226	225	247	248	222	221	251	252	218	217	255
215	257	258	212	211	261	262	208	207	265	266	204	203	269	270	200	199	273	274	196
195	277	278	192	191	281	282	188	187	285	286	184	183	289	290	180	179	293	294	176
296	174	173	299	300	170	169	303	304	166	165	307	308	162	161	311	312	158	157	315
316	154	153	319	320	150	149	323	324	146	145	327	328	142	141	331	332	138	137	335
135	337	338	132	131	341	342	128	127	345	346	124	123	349	350	120	119	353	354	116
115	357	358	112	111	361	362	108	107	365	366	104	103	369	370	100	99	373	374	96
376	94	93	379	380	90	89	383	384	86	85	387	388	82	81	391	392	78	77	395
396	74	73	399	400	70	69	403	404	66	65	407	408	62	61	411	412	58	57	415
55	417	418	52	51	421	422	48	47	425	426	44	43	429	430	40	39	433	434	36

Magic Constant = 4710 Magic Sum =94200

The Fourth Dimension

440	42	43	437	436	46	47	433	432	50	51	429	428	54	55	425	424	58	59	421
61	419	418	64	65	415	414	68	69	411	410	72	73	407	406	76	77	403	402	80
81	399	398	84	85	395	394	88	89	391	390	92	93	387	386	96	97	383	382	100
380	102	103	377	376	106	107	373	372	110	111	369	368	114	115	365	364	118	119	361
360	122	123	357	356	126	127	353	352	130	131	349	348	134	135	345	344	138	139	341
141	339	338	144	145	335	334	148	149	331	330	152	153	327	326	156	157	323	322	160
161	319	318	164	165	315	314	168	169	311	310	172	173	307	306	176	177	303	302	180
300	182	183	297	296	186	187	293	292	190	191	289	288	194	195	285	284	198	199	281
280	202	203	277	276	206	207	273	272	210	211	269	268	214	215	265	264	218	219	261
221	259	258	224	225	255	254	228	229	251	250	232	233	247	246	236	237	243	242	240
241	239	238	244	245	235	234	248	249	231	230	252	253	227	226	256	257	223	222	260
220	262	263	217	216	266	267	213	212	270	271	209	208	274	275	205	204	278	279	201
200	282	283	197	196	286	287	193	192	290	291	189	188	294	295	185	184	298	299	181
301	179	178	304	305	175	174	308	309	171	170	312	313	167	166	316	317	163	162	320
321	159	158	324	325	155	154	328	329	151	150	332	333	147	146	336	337	143	142	340
140	342	343	137	136	346	347	133	132	350	351	129	128	354	355	125	124	358	359	121
120	362	363	117	116	366	367	113	112	370	371	109	108	374	375	105	104	378	379	101
381	99	98	384	385	95	94	388	389	91	90	392	393	87	86	396	397	83	82	400
401	79	78	404	405	75	74	408	409	71	70	412	413	67	66	416	417	63	62	420
60	422	423	57	56	426	427	53	52	430	431	49	48	434	435	45	44	438	439	41

Magic Constant =4810 Magic Sum =96200

The Fifth Dimension

445	47	48	442	441	51	52	438	437	55	56	434	433	59	60	430	429	63	64	426
66	424	423	69	70	420	419	73	74	416	415	77	78	412	411	81	82	408	407	85
86	404	403	89	90	400	399	93	94	396	395	97	98	392	391	101	102	388	387	105
385	107	108	382	381	111	112	378	377	115	116	374	373	119	120	370	369	123	124	366
365	127	128	362	361	131	132	358	357	135	136	354	353	139	140	350	349	143	144	346

146	344	343	149	150	340	339	153	154	336	335	157	158	332	331	161	162	328	327	165
166	324	323	169	170	320	319	173	174	316	315	177	178	312	311	181	182	308	307	185
305	187	188	302	301	191	192	298	297	195	196	294	293	199	200	290	289	203	204	286
285	207	208	282	281	211	212	278	277	215	216	274	273	219	220	270	269	223	224	266
226	264	263	229	230	260	259	233	234	256	255	237	238	252	251	241	242	248	247	245
246	244	243	249	250	240	239	253	254	236	235	257	258	232	231	261	262	228	227	265
225	267	268	222	221	271	272	218	217	275	276	214	213	279	280	210	209	283	284	206
205	287	288	202	201	291	292	198	197	295	296	194	193	299	300	190	189	303	304	186
306	184	183	309	310	180	179	313	314	176	175	317	318	172	171	321	322	168	167	325
326	164	163	329	330	160	159	333	334	156	155	337	338	152	151	341	342	148	147	345
145	347	348	142	141	351	352	138	137	355	356	134	133	359	360	130	129	363	364	126
125	367	368	122	121	371	372	118	117	375	376	114	113	379	380	110	109	383	384	106
386	104	103	389	390	100	99	393	394	96	95	397	398	92	91	401	402	88	87	405
406	84	83	409	410	80	79	413	414	76	75	417	418	72	71	421	422	68	67	425
65	427	428	62	61	431	432	58	57	435	436	54	53	439	440	50	49	443	444	46

Magic Constant =4910

Magic Sum =98200

The Sixth Dimension

445	47	48	442	441	51	52	438	437	55	56	434	433	59	60	430	429	63	64	426
66	424	423	69	70	420	419	73	74	416	415	77	78	412	411	81	82	408	407	85
86	404	403	89	90	400	399	93	94	396	395	97	98	392	391	101	102	388	387	105
385	107	108	382	381	111	112	378	377	115	116	374	373	119	120	370	369	123	124	366
365	127	128	362	361	131	132	358	357	135	136	354	353	139	140	350	349	143	144	346
146	344	343	149	150	340	339	153	154	336	335	157	158	332	331	161	162	328	327	165
166	324	323	169	170	320	319	173	174	316	315	177	178	312	311	181	182	308	307	185
305	187	188	302	301	191	192	298	297	195	196	294	293	199	200	290	289	203	204	286
285	207	208	282	281	211	212	278	277	215	216	274	273	219	220	270	269	223	224	266
226	264	263	229	230	260	259	233	234	256	255	237	238	252	251	241	242	248	247	245
246	244	243	249	250	240	239	253	254	236	235	257	258	232	231	261	262	228	227	265
225	267	268	222	221	271	272	218	217	275	276	214	213	279	280	210	209	283	284	206
205	287	288	202	201	291	292	198	197	295	296	194	193	299	300	190	189	303	304	186
306	184	183	309	310	180	179	313	314	176	175	317	318	172	171	321	322	168	167	325
326	164	163	329	330	160	159	333	334	156	155	337	338	152	151	341	342	148	147	345
145	347	348	142	141	351	352	138	137	355	356	134	133	359	360	130	129	363	364	126
125	367	368	122	121	371	372	118	117	375	376	114	113	379	380	110	109	383	384	106
386	104	103	389	390	100	99	393	394	96	95	397	398	92	91	401	402	88	87	405
406	84	83	409	410	80	79	413	414	76	75	417	418	72	71	421	422	68	67	425
65	427	428	62	61	431	432	58	57	435	436	54	53	439	440	50	49	443	444	46

Magic Constant =5010

Magic Sum =100200

- ❖ Magic Constant of First Dimension is = 4510      Magic Constant of Sixth dimension = 5010  
The Summation of Magic Constant is = **9520**
- ❖ Magic Sum of First Dimension is = 90200      Magic Sum of Sixth dimension = 100200  
The Summation of Magic Sum is = **(190400)**
- ❖ Magic Constant of Second Dimension is = 4610      Magic Constant of Fifth dimension = 4910  
The Summation of Magic Constant is = **9520**
- ❖ Magic Sum of Second Dimension is = 92200      Magic Sum of Fifth dimension = 98200  
The Summation of Magic Sum is = **(190400)**
- ❖ Magic Constant of Third Dimension is = 4710      Magic Constant of Fourth dimension = 4810  
The Summation of Magic Constant is = **9520**
- ❖ Magic Sum of Third Dimension is = 94200      Magic Sum of Fourth dimension = 96200  
The Summation of Magic Sum is = **(190400)**

Select the face of Cube Number...

5



**The Fifth Dimension is:**

**Magic Constant =4910**

**Magic Sum =98200**

- **The Encryption Process ...**

Enter the Plaintext Message

M= "1988"

CipherText (C) = M \* (K=MS) mod P = 1988 \* 98200 mod 1999 = **1259**

**The Encrypted message is: 1259**

- **The Decryption Process ...**

Plaintext= C \* (K<sup>-1</sup>=MS) mod P = 1259 \* 98200<sup>-1</sup> mod 1999 = **1988**

**The Decrypted message: 1988**

- **The Signature Algorithm**

Message Digest

**232** 141 208 5 94 232 134 105 155 187 183 127 242 44 193 22 102 97 27 180 74 167 125 209 22  
111 242 38 108 195 60 195 51 55 117 83 59 190 228 73 157 211 62 235 186 171 186 173 213 86 98  
32 6 99 62 230 104 142 228 69 85 90 167 115

Message abstract for the Message Digest includes also the first byte for easy calculation in tracking and evidence.

**232**

The Signature Process

Sign=Message digest \* Magic Constant mod P

Sign=232\*4910 mod 1999 = **1689**

The Signature is: **1689**

Verify = Sign \* Inverse Magic Constant mod P

Verify = 1689 \* 4910<sup>-1</sup> mod 1999 = **232**

The Verifying is: **232**

**Second Method:** the second method of this model considers the MS and MC as the two keys K1 & K2 respectively (MS=K1, MC=K2)

Magic Sum =100200 = K1, Magic Constant =5010= K2.

- **The Encryption Process ...**

C=K1 \* M + K2 mod P.

=98200 \* 1988 +4910 mod 1999 =172.

- **The Decryption Process ...**

M=K1<sup>-1</sup> (C-K2) mod P.

570(172 - 4910)

=98040 - 2798700 =-2700660 mod 1999

2700649-2700660 = -11 mod 1999= 1988 =M

- **The Signature Algorithm**

- **Sing Process**

C=K1 \* M + K2 mod P.

=98200 \* 232 +4910 mod 1999 =709.

- **Verifying Process**

M=K1<sup>-1</sup> (C-K2) mod P.

570(709 - 4910)

=404130 - 2798700 = -2394570 mod 1999

1767- mod 1999= 232 =M

## X. MOTIVATIONS AND LIMITATIONS FOR PROPOSED PUBLIC KEY

There are several motivations to design the proposed public key model, since most of the public key methods mainly depend upon the DLP and IFP and these methods are highly time consuming. The core idea for the proposed model concentrates on the design, a fast and strong method that is based on the magic cube mathematical problem, in order to guarantee the complete security and invulnerability against the malicious attacks. The complexity of this cipher includes that the eavesdroppers should try all possible probabilities of construction the magic cube' matrices, starting from the unknown random value that acts as a starting value for the construction and unknown dimension value which needs a lot of estimation and guessing. This scheme allows fast encryption and decryption process in addition to the fast signature generation and verification, as well as it enlarges the search space against the brute force attack and consequently increases the complexity. There are several negatives and limitations that are considered one of the main problems and the limited side in the public-key cryptography which involves how to evidence that the asymmetric key is authoritative and it has not altered and changed by another key or by an intruder or even an unknown intercepted person. Perhaps the most vulnerable attack on the public key cipher is the man-in-the-middle attack, in which a third malicious party impersonates the personality of the authorized person by the intercepting and modifying the public key. An active adversary in the middle communication manipulates and modifies the messages and the implication deceives the two communicated parties [17]. In order to agree on a key which is exclusively shared between Alice and Bob, these principals must make sure that the messages they receive in a protocol run are indeed from the intended principals. A trusted third party may be used to represent as a certificate authority that is verifying from the identity of persons using the authentication system. The public key in general terms or asymmetric cryptosystem compared with the symmetric cryptosystem take much more time in the established key for encryption and decryption processes since it uses sophisticated mathematical problems in its construction. The random choices for some numbers to construct the magic cube whence starting value, difference value and the fixed multiplied value to generate the private key give a more resistant against the attacks. Unacceptable selections for these random numbers represent a basic restriction and may open the door in front of the passive and active attacks. So, one should be careful to ensure that the identity of the parties takes place by using a trusted third party or mediated party. Eventually, the main purpose for the extended magic square to magic cube is to treat with six parties or terminals on network instead of one using the same field and the same module number according to the DH algorithm in order to exchange the secret key. Hence, each face from the cube can be represented as an independent terminal. Thus, the main purpose for the generalized magic cube is to create a simple method based on the folding concept that constructed the magic cube easily. The magic cube gives more potential for the selection of ciphering key through the generation process as each face or dimension in the magic cube can be used as a key generation method for the encryption/ decryption process and the signature algorithm.

## XI. ANALYSIS AND EXPERIMENTAL RESULTS

The security of the proposed cipher based on the mixed more than one mathematical problem to apply high margin of security. An efficient, secure and fast algorithm employed to applies secure digital communication which is based on the hardness of some problems in number theory. The magic square construction also based on various

techniques that give more strength to defeats the attacks and increases the probability of resistant in front the statistical analysis. The efficiency of a proposed cipher is relied on the time elapsed for encryption/decryption and the way it produces different cipher-text from a plaintext. With respect to efficiency, as it well known that the most of the public-key cipher suffers from the difficulty of the key generation and the parameters selection for the session establishment and the key agreement. The proposed cipher on the software platforms offers a cost effective and flexible solution for the key exchange (key agility) and encryption/decryption. The adoption of the magic cube mathematical problem could significantly change the nature of public key cryptography and the manner through which will be treated, in addition to the behavior and the style of attacks. We have introduced a simple comparison among three different public key ciphers in the below as explained in Figure 8. which illustrates the implementation of run time in seconds to achieve the encryption and decryption operation for the three messages with different size (1000 char, 2000 char, and 3000 char) respectively. Figure 9. represents the running time of the signature and verification algorithms for the same message. In this test there is no need to take different messages lengths, because the execution time will be based on the message digest of the original message.

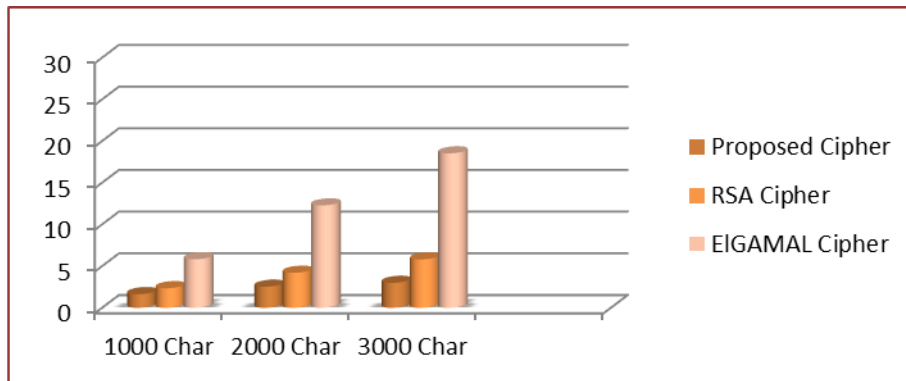


Figure 8. Encryption and Decryption Chart's Time for Different Algorithms

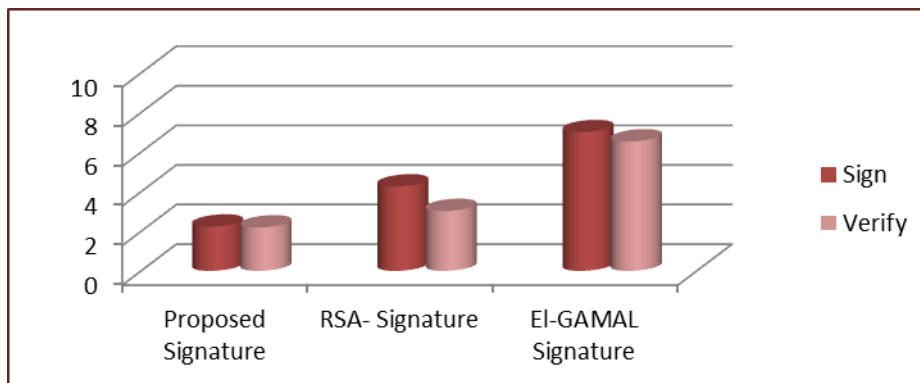


Figure 9. Signature and Verification Chart's Time Conclusions for Different Algorithms

We have shown that the proposed model give a good insight and introduced a smart method in the designing processes that paved the way front the new mathematical comprehension which related to the probability of dimension for the magic cube construction. Since the search space and the complexity increased dramatically with the increasing dimension. The basic idea described in this paper focused on the clue of creates a confidential

communication channel with a secret sharing between the communicating parties in the presence of malicious adversaries. The magic cube mathematical problem has been exploited and played a vital role in encryption/decryption and signing/verifying operations. It gives a remarkable significant speed and reduced the costs as well as improves the efficiency and security margin.

## XII. CONCLUSION

In the present study we have developed a new model of asymmetric cipher that comprises the improved technique for the public key by depending upon the magic square and magic cube techniques. So, the proposed model gives a good insight and introduces a smart method in the designing processes that paved the way for the new mathematical comprehension which is related to the probability of dimension for the magic square & magic cube construction, since the research space and the complexity of magic cube increases dramatically with the increasing dimension. The basic idea is focused on the clue of creating a confidential communication channel with a secret sharing between the communicating parties in the presence of malicious adversaries. The magic cube mathematical problem has been exploited, and it plays a vital role in encryption/decryption and signing/verifying operations with two different methods. It gives a remarkable significant speed and reduces costs as well as improvement in the efficiency and security margin. The proposed of the folded cube method is considered as the simplest and nearly the fastest method to construct the magic cube, since it is based on the folded procedures and the traditional magic square methods that can be constructed with any order easily. There is no existence for any real difficulty in the construction of any cube with this technique, because it based on the folded process for the magic square methods, and does not need a strong mathematical comprehension or experience in the geometrical aspects.

## ACKNOWLEDGMENT

The authors would like to thank Dr. Hazim H. Muarich for his great efforts and they are grateful for his valuable comments and suggestions from a linguistic point of view.

## REFERENCES

- [1] Zhao-Xue Chen and Sheng-Dong Nie, "Two Efficient Edge Detecting Operators Derived from 3 X 3 Magic Squares", Proceedings of the International Conference on Wavelet Analysis and Pattern Recognition, Beijing, China, 2-4 Nov. -4244-1066-5/07/\$25.00 ©2007 IEEE.
- [2] GUOCE XIN, "Constructing All Magic Squares of Order Three", September arXiv:math/0409468v1, math.CO, 24 Sep 2004.
- [3] D. Rajavel and S. P. Shantharajah, "Cubical Key Generation and Encryption Algorithm Based on Hybrid Cube's Rotation", Proceedings of the International Conference on Pattern Recognition, Informatics and Medical Engineering, 978-1-4673-1039-0/12/\$31.00 ©, March 21-23, 2012 IEEE.
- [4] Adam Rogers and Peter Loly, "The Inertia Tensor of a Magic Cube", American Association of Physics Teachers, Am. J. Phys. © June 2004. <http://aapt.org/ajp>.
- [5] Brendan Lucier, "Unfolding and Reconstructing Polyhedra", A thesis Presented to the University of Waterloo in fulfillment of the thesis requirement for the degree of Master of Mathematics in Computer Science, c Brendan Lucier, 2006.
- [6] Nitin Pandey and D.B.Ojha, "Secure Communication Scheme with Magic Square", Volume 3, No. 12, December 2012 Journal of Global Research in Computer Science.
- [7] A. Dharini, R.M. Saranya Devi, and I. Chandrasekar, "Data Security for Cloud Computing Using RSA with Magic Square Algorithm", International Journal of Innovation and Scientific Research, ISSN 2351-8014 Vol. 11 No. 2 Nov. 2014, pp. 439-444 © 2014 Innovative Space of Scientific Research Journals <http://www.ijisr.issr-journals.org/>.

- [8] Gopinath Ganapathy, and K. Mani, "Add-On Security Model for Public-Key Cryptosystem Based on Magic Square Implementation", Proceedings of the World Congress on Engineering and Computer Science 2009 Vol I WCECS 2009, October 20-22, 2009, San Francisco, USA.
- [9] D.I. George, J.Sai Geetha and K.Mani, "Add-on Security Level for Public Key Cryptosystem using Magic Rectangle with Column/Row Shifting", International Journal of Computer Applications (0975 – 8887) Volume 96– No.14, June 2014.
- [10] Paul Carus, "The Zen of Magic Squares, Circles, and Stars", Copyright © 2002 by Clifford A. Pickover.
- [11] Andrews, W. S. Magic Squares and Cubes. Chicago: Second Edition. Revised and Enlarged Open Court Publishing, 1917.
- [12] H. D. Heinz & J. R. Hendricks, "Magic Squares Lexicon: Illustrated", Copyright © 2000 by Harvey D. Heinz
- [13] Steve Burnett and Stephen Paine, "RSA Security's Official Guide to Cryptography", Copyright © 2001 by The McGraw--Hill Companies.
- [14] Serge Vaudenay, "A Classical Introduction to Cryptography Applications for Communications Security", Swiss Federal Institute of Technologies (EPFL), @ 2006 Springer Science Business Media, Inc.
- [15] Zhao-Xue Chen and Sheng-Dong Nie, "Two Efficient Edge Detecting Operators Derived From 3X3 Magic Squares", Proceedings of the International Conference on Wavelet Analysis and Pattern Recognition, Beijing, China, 2-4 Nov-4244-1066-5/07/\$25.00 ©2007 IEEE.
- [16] Guoce Xin, "Constructing all Magic Squares of Order Three", Discrete Mathematics, 0012-365X/\$ - see front matter © 2007 Published by Elsevier B.V, pp3393-3398.
- [17] Karim Sultan and Umar Ruhi, "Overcoming Barriers to Client-Side Digital Certificate Adoption", (IJCSIS) International Journal of Computer Science and Information Security, Vol. 13, No. 8, August 2015.

#### Authors' Profiles



**Omar Abdulrahman Dawood** was born in Habanyah, Anbar, Iraq (1986), now he lives in Ramadi, Anbar. He obtained B.Sc. (2008), M.Sc. (2011) in Computer Science from the College of Computer, Anbar University, Iraq. He was ranking the first during his B.Sc. and M.Sc. studies. He is a teaching staff member in the English Department in College of Education for Humanities, Anbar University, and currently he is a Ph.D. student at the Technology University- Baghdad. His research interests are: Data and Network Security, Coding, Number Theory and Cryptography.



**Prof. Abdul Monem S. Rahma** Ph.D Awarded his M.Sc. from Brunel University and his Ph.D. from Loughborough University of technology United Kingdom in 1982, 1984 respectively. He taught at Baghdad university Department of Computer Science and the Military Collage of Engineering, Computer Engineering Department from 1986 till 2003. He holds the position of Dean Asst. of the scientific affairs and works as a professor at the University of Technology computer Science Department. He published 88 Papers, 4 Books in the field of computer science, supervised 28 Ph.D. and 57 M.Sc. students. His research interests include Computer Graphics Image Processing, Biometrics and Computer Security. And he has attended and submitted in many scientific global conferences in Iraq and many other countries. From 2013 to Jan. 2015 he holds the position of Dean of the Computer Science College at the University of Technology.



**Abdul Mohssen J. Abdul Hossen** is an Associate Professor of Applied mathematics, Computer Science Department, University of Technology, where he teaches undergraduate and graduate courses in Mathematics. Dr. Abdul Hossen received the B.Sc. in Mathematics from Mustansiriyah University, Iraq 1977, the M.Sc. degree in Applied Mathematics from Bagdad University, Iraq. in1980, the Ph.D. in Applied Mathematics from University of Technology, Iraq, 2005. He is a member of the IEEE system, and Member of the Editorial Journal.

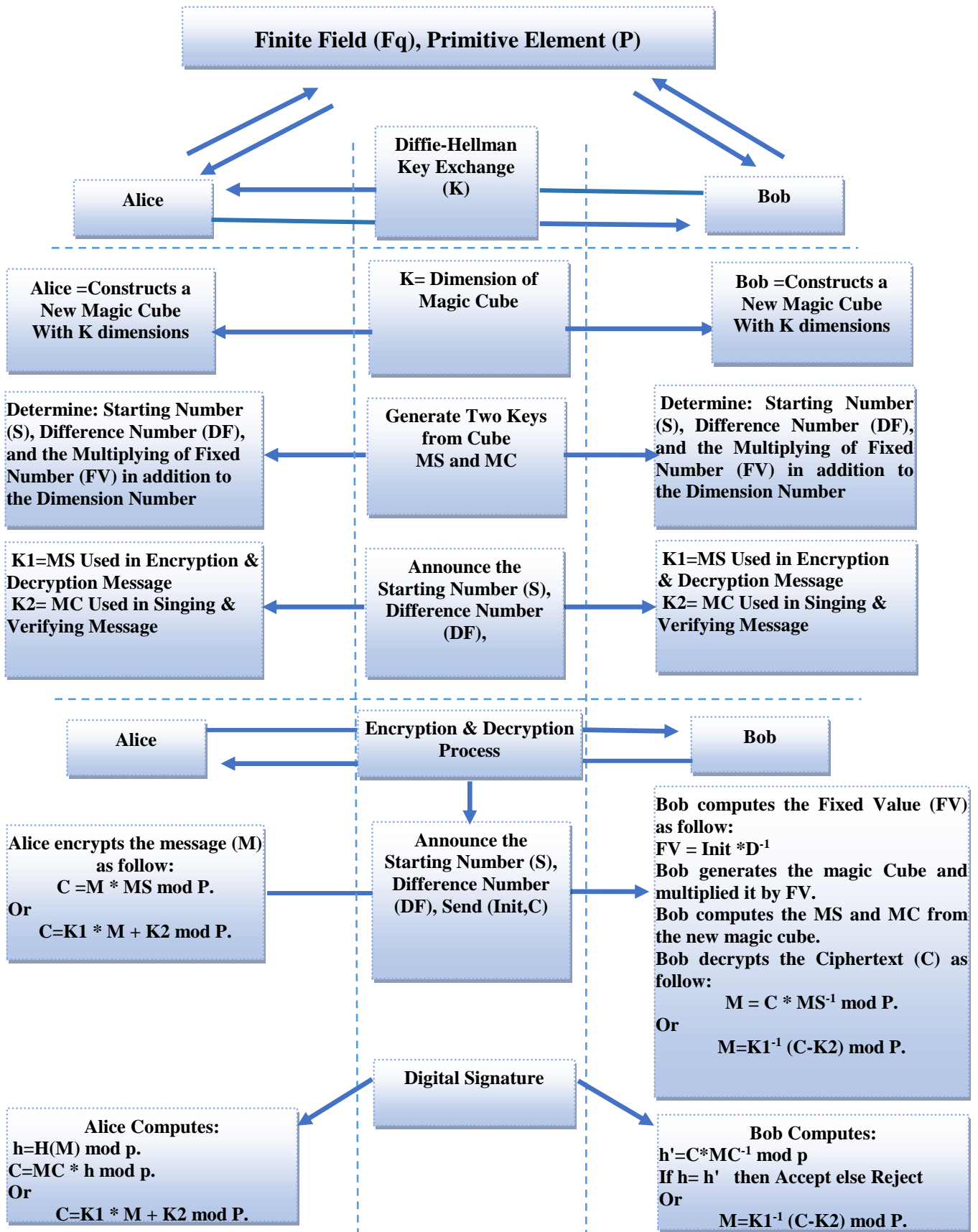


Figure 7. The Proposed model



# Defining Project Based Learning steps and evaluation method for software engineering students

Mohammad Sepahkar  
Department of Computer Engineering  
Islamic Azad University of Najafabad, Iran

Faramarz Hendessi  
Department of Computer & Electronic  
Isfahan University of Technology, Iran

Akbar Nabiollahi  
Department of Computer Engineering  
Islamic Azad University of Najafabad, Iran

## ABSTRACT

**Needing well educated and skillful workforce is one of the top items in industrial top priority list. But traditional education systems only focus on teaching theoretical knowledge to students which leads to lack of practical experience in them. Therefore modern pedagogy came to overcome this problem. Project based learning is one of these interactive learning pedagogies which is mostly used in engineering educations all over the world. In this research, we review a case study of executing a project based learning program in Isfahan University of Technology, Computer Engineering Department. During this overview, we explain all the steps needed for holding a PjBL curriculum with subject of software development. Finally we discuss about evaluation method for Project based learning programs.**

Project based Learning, Education Pedagogy, Traditional Pedagogy, Software development, Team setup, Evaluation

## INTRODUCTION

Due to the economic situation in last decades, and unemployment rate, there is a very high competition between workforces to be employed in companies which need well educated and skillful employees. Thus, the role of education system in preparing them is extremely bold. In other word, it could be said that "the primary purpose of higher education, in all essence, is to prepare students for the workplace" [1]. But traditional educational methods only provide theoretical, technical and fundamental knowledge of engineering [2] which are not enough for employment; and companies inevitable to expenditure for them to prepare to get the job.

Because of this weakness in traditional pedagogy, modern education pedagogy was born which is based on active learning and encourages students to be active participants learning process [3]. These modern pedagogies include "Problem-Based Learning (PBL)", "Cooperative & Collaborative Learning", "Project-Based Learning (PjBL)";

Problem-Based Learning (PBL) is defined as "the learning which results from the process of working towards understanding of, or resolution of a problem" [4]. PBL has been described in medical field since early 1960s [5]. The purpose of problem based learning is (1) Acquisition of knowledge that can be retrieved and used in a professional setting. (2) Acquisition of skills to extend and improve one's own knowledge. (3) Acquisition of professional problem-solving skills [6]. It is a learner-centered approach. In this learning process some unstructured problems are used as the starting point and anchor [7]

Cooperative & Collaborative Learning: cooperative learning is highly structured and includes positive interdependence (all members must work together to complete a task) as well as individual and group

accountability [8]. Collaborative learning needs not be as structured as cooperative learning and may not include all of its features. Individual accountability is not a necessary element of collaborative learning [9]

Project-Based Learning (PjBL) is a teaching method that involves students in learning required ability. Student-influenced inquiry process structured around complex, authentic questions and carefully designed products and tasks [10], [11]. Students' interest, critical thinking skills, relationship ability, and team working skills, were improved when they worked on a PjBL activity [12] and surely these skills are unable to be developed by solely depending on traditional methods [13]. In other words we can say PjBL means learning through experience [14]. It has been proven that through PjBL, students' generic skills can be improved too [15]. Student can learn time and resource management, task and role differentiation, and self-direction, during these projects. [16]

PjBL projects are central pedagogy. PjBL is not peripheral to the curriculum. It also focused on questions that guides students to face main. PjBL involves students in a beneficial experiment so it is student-driven pedagogy. Those projects, which they involved, are realistic, not school-like. [17] Student learning in this method is inherently valuable because it's practical and involves some skills such as collaboration and reflection. [14] The main objective of PjBL is development and improvement of technical and non-technical skills and it provides real engineering practice for students [13].

Modern development of computer equipment and information technology (IT) calls for new education and adequate training. [18] So we can say PjBL is very important pedagogy for IT and computer engineering high level educations. Based on Gallup organization report, on average, 9 in 10 students said that study programs should include communication skills, teamwork, and learning to learn techniques [19] which are also great factors in IT educations.

Using PjBL in any field of study and in any level of education, may have some difference. In this research we are going to establish a method for PjBL in software engineering bachelor education by investigating a case study of PjBL in Isfahan University of Technology, computer engineering department. We review this case study and during that, we define a method for holding similar curriculums. In each part of this review we explain what is needed for running a PjBL program.

#### PROJECT-BASED LEARNING PROJECT VS NORMAL PROJECT

The first part of PjBL is defining a real project. But there are some major differences between project based learning software development project doing and normal one, that must completely be understood such as the ones listed below:

- Project manager (teacher) is not only responsible for tasks such as scheduling, resource management and etc, but also engaged in educating students and improving their practical skills. From other perspective, teacher does not teach students in detail the way of doing things, as they must work in group to complete the task given to them. Instead, he guides the students in order to make sure they are on the right path. [13]
- Unlike a normal project, there is not a dedicated role for each member. Each participant plays different role in order to have some new experience in that category. Even teacher, as was told before, plays different roles like tutor, coach or facilitator. [20]
- Although in a normal project, project manager is responsible for project success, but in project based learning project there is more pressure on him. Indeed he is responsible for each task that is done by any of participants. Therefore if he has not enough experience about project subject, failure of the project will be certain. In other words theoretical knowledge is not sufficient for the teacher and he must have practical experience in that field. [21]
- In normal project, time is an important factor. So each task must be done in specific time. But in project based learning, some tasks will be done several times in different ways so that students discover the better way to solve a problem. It is also possible that teacher lets students experience a wrong way, and see the result so that they avoid the mistake in similar conditions. On the other hand we can say project based learning is an "outlet for every student to experience success" [22] by themselves.

#### SETUP A PROGRAM

To setup a project based learning program, we must define its specification. This program specification has been listed in Table 1; Students, who are involved in this program, participated in that as their internships course (one of undergraduate lessons in IUT)

Table 1 : Project base learning program specifications

Number of participants	22 students
Curriculum timing	2007 Summer
Participant educational degree	Undergraduate students
Field of Study	Computer engineering
Project subject	Educational web system
Programming language	ASP .Net (C#) or php
Database	SQLServer or MySQL

All steps of this program is shown in Figure 1



Figure 1 : Steps of a software project based learning program

### Defining Project

Project definition is the major part of project based learning program. Research shows that “poor project definition is recognized by industry practitioners as one of the leading causes of project failure.” [23]

Selected project must have special conditions. This project must “have sufficient potential for exploration and investigation, allow for the opportunity of problem-solving, collaboration, and cooperation, and provide the opportunity of construction.” [24] Also it shouldn't have a specific highly sensitive deadline, but rather an open-end scheduling. In order to do this, we should look for a taskmaster with high flexibility. So we selected "Electronical Education Office (EEO) in Isfahan University of Technology (IUT) IT Center" as our taskmaster for this program.

EEO was responsible for ICDL education program for employers of government organizations. Because of high number of these educational programs, they needed a web-based program by which participants could register and follow up their educational situations, points and etc. So a project with the subject of "Define a web-based Educational Program" was defined as our project for this curriculum.

### Choosing Software Development Environment

Since the main goal of this program was education of participants, practical experience in web programming, without any preference for a specific programming language, choosing a unique environment did not appear to be necessary. Thus due to students' interest, it was decided to use two development environments for the project. One with ASP.Net (C#) and SQL Server as database engine, and another with php and MySQL as database engine as shown in Table 2.

Table 2 : Developing Inviroments

Team	Programming Language	DBMS
A	ASP.Net C#	SQL Server
B	php	My SQL

### Team Setup

The traditional hierarchical model of leadership is outdated. These days flatter industrial models where leadership is shared amongst the various individuals in a team, is widely used [25], so defining a team structure is a very important task that must be done before curriculum is started. Although the students are mostly unskilled but they must define the responsibilities of team members in team setup, and there is some success history of this work [26]. Team structure that is defined for this purpose prefers not to be hierarchical as common project teams. Breaking down team leadership to flat management model, leads to rotation of leadership and share of power. [25]

In this case study, according to development environment, all participants were divided in two teams based on their basic knowledge and interest. Twelve students in group A (ASP.Net C# & SQL Server) and ten students in group B (php & MySQL).

Due to skills and abilities some roles were defined in each group, as shown in Table 3.

Table 3 : Roles in Groups

Role	Count	Descriptions
Teacher	1	A Person with teaching and project management experiments in similar projects. He also must be experienced in both project development environments. He is responsible for project management, scheduling, student guiding and etc.
Teacher Assistant	1	A person with teaching and managing students, experiments and technical knowledge in both development environments is selected as TA. He is responsible for coordination of teams and guiding students during project phases.

Role	Count	Descriptions
<b>Team Headman</b>	1 in team A 1 in Team B	In each team, a person who has enough ability for managing team and communication with team members, and also has technical knowledge about DBMS and programming language which are selected for that team, is selected as TH.
<b>System Analyze Headman</b>	1	Since analyze and requirement engineering of system is same for both teams, one person is sufficient. Usually this role is assigned to teacher assistant who has enough experiment in system analyze.
<b>Database Design Chief</b>	1 in team A 1 in Team B	In each team, the most skillful person in database designing, is selected as DDC. This person should have enough ability of designing tables and coding needed store procedures and functions in that DBPMS.
<b>Database Design Subteam Members</b>	2 in team A 2 in Team B	These students should have primary familiarity with that DBMS. They design and implement database with the help of DDC.
<b>Programming Subteam Chief</b>	1 in team A 1 in Team B	In each team, the most skillful person in programming with their programming language, is selected as PSC. This person should have enough ability in coding with that programming language.
<b>Programming Subteam Members</b>	6 in team A 5 in Team B	These students should have basic familiarity with that programming language. They write codes of programs with the help of PSC.
<b>Test Subteam Chief</b>	1	He must have enough mastery in the whole system functionality. He is responsible for designing test scenario. Teacher assistant can act as TSC.
<b>Test Subteam Members</b>	1 in team A 1 in Team B	They must execute all test cases that are designed by TSC.

Due to the goal of this program, which is teaching practical skills to students, unlike real software development teams for software manufacturing, in which each person works in a special field, students who participate in this program, have not a fixed role and they may play different roles in various teams. For example in analyze phase both team members act as analyze team members to obtain real project analyze experiments. Also since test subteam members, cannot start their work before programming subteam members start development, they can be involved in programming.

Like teacher, teacher assistant is not a student, and he has experiment and knowledge about roles that he plays during the project. As system analyzer responsible, he is initiator of system analyze. He also is responsible for testing system.

In a project based learning program, teacher acts as project manager. He plays different roles during the project development. In addition to the duties of project manager including scheduling, supervise execution, coordinating team members and ensuring project success. Another main duty of the teacher, is educating team members. He must guide students during each phase of the project, while they perform their duties as well; they also acquire needed skills in that field.

#### Holding Workshop

In case that participants do not have enough basic experiment, during an intensive workshop, required basic information will be transferred to them.

In this case study, the majority of participants did not have the basic knowledge for doing this project. Thus we held a ten days workshop for teaching basic information about the subject of the project. At the end of this workshop, all participants had the basic knowledge for developing web-based programs.

#### System Analyze

The first phase of the project is system analyze. The main purpose of this phase is gathering required information and requirement engineering. Although the output of this phase is used as the input for the other phase, its accuracy is very important. Therefore, teacher assistant is directly responsible for this phase.

During this phase, all students involve system analyze. If it is possible to have meeting with stakeholders of the project, students can participate in that, but every question about system analyze, especially for people who do not have enough knowledge in that field, must be asked under the supervision of system analyze headman or teacher because asking basic or irrelevant questions, may cause suspicion of stakeholders. In this situation the person who is questioned does not answer them correctly, which finally leads to project failure. So students usually attend in meetings as an observer and in few occasions, with assistance of system analyze headman or teacher, can ask their questions from stakeholders directly.

After first meeting with stakeholders, second meeting with students and teacher will be held. In that session, teacher plays stakeholders role and answers all questions. Teacher assistant guides students to ask correct questions, and asking any question is allowed there. If any question is asked in that session, which teacher cannot answer, in the next meeting with stakeholders it will be asked.

Analyze meetings will continue until system analyze is finished and designing phase can be started.

### System Design

In software development project the first step of system designing is database design. Database should be designed according to the requirements that achieve in analyze phase. This task, independent of tools, and with supervision of teacher and support of teacher assistant will be done.

In some meeting that all students are present, database designing is completed. Then teacher asks some smart question about technical weaknesses of the first design, and students find answers to them and correct database design with the help of teacher assistant.

Program structure design is in progress in parallel by participants. Teacher asks some questions about how this design covers all requirements that was derived in previous phase, and students find suitable answers and correct initial design with the help of teacher assistant. It may be prepare a prototype of the software interface to explain what it is supposed to be achieved better, by each programming subteam chief.

At end of this phase, database and program design is finalize and confirmed by teacher, and next phase will begin.

### Implementation

Henceforth, each team operations preforms and monitors independently. Each team starts implementation according to the design which was prepared in the previous phase. Database design chief, with subteam members and under teacher supervision implement database in the associated DBMS. If any problem has occurred, at first they ask teacher assistant for help and if problem persisted, they call teacher to resolve that. Finally the complete database and its related functions are implemented and some unreal data will be inserted into it for testing purpose.

Coding operations are also performed in parallel, under supervision of teacher and supporting of teacher assistant, based on the design which was achieved in the previous phase. In case of difficulty, students can call teacher assistant and also teacher for help.

### System Test

After start of coding, testing operations will begin. Testing operations will be performed for each implemented module and the whole system. Designing test scenario will be done by test subteam chief (who is teacher assistant) with cooperation of test subteam members.

After defining each scenario, execution operation will begin by test subteam members. If any error is found, error list is sent to programming subteam to be addressed. If needed, teacher assistant provides necessary recommendations for fixing them and avoiding recurrence of similar errors.

After finishing implementation of final system, test will be executed under direct supervision of teacher to ensure that no problem is ignored.

## EVALUATION

At the end of the project based learning program, we must have an evaluation to measure the success rate. Project based learning program success factors include:

- Project success
- Student practical skills improvement

By measuring the above criteria we prove that our case study is completely successful.

### Project success

Since the final artifacts demonstrate the capability of students that participate in this curriculum, we can use it to determine the degree of program success. On the other hand we can say that project success or failure, is the first indicator that shows us if our education program has been successful or not. For project success we need a team with necessary experience in each project team. But students who participate in this program, usually do not have the basic skills for doing a real project. So if teacher (as project manager) can finish the project successfully with such inexperienced people, this means success of the curriculum, because participants can obtain necessary skills to do a real project.

In this case study, both projects which were developed by students were completely successful. And finally the project that was developed in .Net environment was deployed for Electronical Education Office and they started using that program for managing registration of applicants for ICDL classes from the next fall semester.

### Students' progress

Students who are included in this program shall move from novices to experts in the domain of knowledge. [27] So the other criteria that helps us to measure the curriculum success rate, direct student questioning about improvement of their practical skills, which called self-evaluations. [25] This evaluation method enables students

to focus on their learning process and allows them to see their progress. [28] Self-evaluation gives students a sense of accomplishment and further instills responsibility for learning [29]

It is obvious that their improvement rate is different according to their first skill levels. For example coding skills improvement for a student who has not any coding experience, is more than a student who has initial experience in that.

For this kind of evaluation usually we can ask a question with five possible answer (point 1 to 5) which determine their progress in curriculum. [30] In this case study, we inquire participants about improvement of their practical skills. The average point of this evaluation is 4.55 from 5. The result of this is shown in

Table 4, which ensures the success in improvement of student's practical skills, certificated by themselves.

Table 4 : Participants inquiry result about improvement in their skills

Bad	Not enough	Good	Very good	Excellent
0 %	0 %	9%	27 %	64%

## CONCLUSION

Like other curriculum, project based learning program needs well planning. Project type influences project based learning program planning. For software development project based learning program we have these steps:

- Project definition: project definition is a very important part of curriculum. The project which is selected for a project based learning program must have special conditions: It should have enough potential to provide an opportunity for participants to increase their practical skills in that subject. Certainly because of its educational nature of this program, the project must not have a critical deadline; therefore it should have a very flexible company as its taskmaster who hasn't a very fixed scheduling for delivering of the project.
- Choosing software development environment: If the company is flexible enough and has not any term to use a specific development environment, we need to select a suitable development environment to develop the project. We must select a popular development environment to improve practical ability of students in that. And if it is possible, we can use more than one environment to develop the project.
- Setting up a team: The most sensitive work in project based learning program is setting up a team. According to the type of project, the team definition is different. For a software development project, a team includes these members:
  - Teacher as project manager who is responsible for educating students and project success.
  - Teacher assistant for helping teacher in educating students and doing the project.
  - Team headman is one of participants with higher level of knowledge in that category.
  - System Analyze Headman who is responsible for system analyzes and usually is the teacher assistant.
  - Database design subteam members who are responsible for database design. The subteam chief is a student who has enough knowledge about designing a database in the DBMS.
  - Programming subteam members who are responsible for coding project. The subteam chief is a student who has some experience in coding with the program language.
  - Test subteam members who are responsible for testing program. The subteam chief is a student who is the most skillful student of the team in the development environment and system analyze.
- Holding a workshop: Typically, participants do not have basic knowledge about the project subject. Thus it is essential to hold a workshop and teach them some necessary basic information before starting the project.
- System analyze: The first and the most important phase of software developing project is analyze. In project based learning analyze phase is a little different. It is directly managed by teacher and teacher assistant and student mostly acts as observer in stakeholder meeting. And there would be some simulation analyze meetings which teacher plays stakeholder roles and students can ask questions.

- System design: In this part, role of students is more bolded. They design system according to the data from the previous phase with teacher assistant support under teacher supervision.
- Implementation: In this phase students are really involved because they are going to prepare the most important project artifact which is coding of the program. They implement the software according to the design data from previous phase, in compliance with standards and patterns that teacher defines for them.
- System test: The last phase of software development in project based learning is test phase. (note deployment and support could be ignored in this curriculum). In this phase which is started a little bit after beginning of implementation phase, students test each module of the project and finally they test the whole of project under teacher supervision.

#### REFERENCES

- [1] NCIHE, "Higher education in the learning society [Report of the National Committee of Inquiry into Higher Education: 'The Dearing,'" Norwich: HMSO Available at [<https://bei.leeds.ac.uk/Partners/NCIHE/>], 1997.
- [2] S. Kumar and J. K. Hsiao, "Engineers learn "soft skills the hard way": planting a seed of leadership in engineering classes," *Leadership and Management in Engineering*, vol. 7, no. 1, pp. 18-23, 2007.
- [3] M. Huang, D. Malicky and S. Lord, "Choosing an Optimal Pedagogy: A Design Approach," in 36th ASEE/IEEE Frontiers in Education Conference, San Diego, CA, October 28 – 31, 2006.
- [4] H. S. Barrows and R. M. Tamblyn, "Problem-based Learning: An Approach to Medical Education," Springer, 1980.
- [5] J. E. Mills and D. F. Treagust, "Engineering education – Is problem-based or project-based learning the answer?," *Australian Journal of Engineering Education*, 2003.
- [6] J. C. Perrenet, P. a. J. Bouhuijs and J. G. M. M. Smits, "The Suitability of Problem-based Learning for Engineering Education: Theory and practice," *Teaching in Higher Education*, vol. 5, no. 3, pp. 345-358, 2000.
- [7] O. Tan, "Problem-based Learning Innovation: Using Problems to Power Learning in the 21st Century," in Thomson Learning, Singapore, 2003.
- [8] K. A. Smith, S. D. Sheppard, D. W. Johnson and R. T. Johnson, "Pedagogies of Engagement: Classroom-Based Practices," *Journal of Engineering Education*, vol. 94, no. 1, pp. 87-101, 2005.
- [9] M. Huang, D. Malicky and S. Lord, "Choosing an Optimal Pedagogy: A Design Approach," in 36th ASEE/IEEE Frontiers in Education Conference, San Diego, CA, October 28 – 31, 2006.
- [10] W. Moylan, "Learning by Project: Developing Essential 21st Century Skills Using Student Team Projects," *International Journal of Learning*, vol. 15, pp. 287-292, 2008.
- [11] T. Markham, J. Larmer and J. Ravitz, *Project based learning handbook: a guide to standards-focused project based learning for middle and high school teachers*, Novato, CA.: Buck Institute For Education, 2003.
- [12] M. Neo and T. K. Neo, "Engaging students in multimedia-mediated constructivist learning - Students' perceptions," *Educational Technology & Society*, vol. 12, no. 2, p. 254-266, 2009.
- [13] M. K. NOORDIN, A. N. M. NASIR, D. F. ALI and M. S. NORDIN, "Problem-Based Learning (PBL) and Project-Based Learning (PjBL) in engineering education: a comparison," in *Proceedings of the IETEC '11 Conference*, Kuala Lumpur, Malaysia, 2011.
- [14] G. Solomon, "Project-Based Learning: a Primer," *Technology and Learning*, vol. 23, no. 6, 2003.
- [15] K. Mohd and e. a. Yusof, "Promoting Problem-Based Learning (PBL) in Engineering Courses at the Universiti Teknologi Malaysia," *Global Journal of Engineering Education*, vol. 9, no. 2, 2005.
- [16] Y. V. Zastavker, M. Ong and L. Page, "Women in Engineering: Exploring the Effects of Project-Based Learning in a First-Year Undergraduate Engineering Program," in *October 28 – 31, 2006, San Diego, CA, 36th ASEE/IEEE Frontiers in Education Conference*.
- [17] J. Thomas, "A review of research on project-based learning,," bobpearlman, 2000. [Online]. Available: [www.bobpearlman.org/BestPractices/PBL\\_Research.pdf](http://www.bobpearlman.org/BestPractices/PBL_Research.pdf). [Accessed 1 Nov 2009].
- [18] S. Grozdev and E. Angelova, "Word Processing as a competence in qualification of teachers of information technologies in Science, Education and Time as our Concern,," in *Proceedings of the Jubilee Scientific Conference with International Participation*, Bulgarian, Nov. 30–Dec. 1, 2007.
- [19] "Students and Higher Education Reform Special Target Survey," The Gallup Organization, 2009.
- [20] "The Comparison of Problem-based Learning (PmBL) Model and Project-based Learning (PtBL) Model," in *International Conference on Engineering Education – ICEE 2007*, Coimbra, Portugal, 2007, September.
- [21] R. W. Marx, P. C. Blumenfeld, J. S. Krajcik and E. Soloway, "Enacting project-based science," *The Elementary School Journal*, vol. 97, no. 4, pp. 341-358, 1997.
- [22] S. Wolk, "PBL: Pursuits with a purpose,," *Educational Leadership*, vol. 52, no. 3, pp. 42-45, 1994.
- [23] C.-S. Cho and E. Gibson, "Building project scope definition using project definition rating index," *Journal of Architectural Engineering*, vol. 7, no. 4, pp. 115-125, 2001.
- [24] Akinoglu and Orhan, "Assessment of the inquiry-based project application in science education upon Turkish science teachers' perspectives," *Education*, vol. 129, no. 2, pp. 202-215, 2008.
- [25] K. Cain and S. Cocco, "Leadership Development through Project Based Learning," in *Proc. 2013 Canadian Engineering Education Association (CEEAI3) Conf.*, Montreal, QC, 2013.



- [26] B. Katja and A. Wiek, "Do We Teach What We Preach? An International Comparison of Problem- and Project-Based Learning Courses in Sustainability," *Sustainability*, vol. 5, no. 4, pp. 1725-1746, 2013.
- [27] M. M. Grant and R. M. Branch, "Project-based learning in a middle school: Tracing abilities," *Journal of Research on Technology in Education*, vol. 38, no. 1, pp. 65-98, 2005.
- [28] A. R. M. Baharuddin, M. D. Khairul Azhar, J. Kamaruzaman and A. G. Nik Azida, "Project Based Learning (PjBL) Practices at Politeknik Kota Bharu, Malaysia," *International Education Studies*, vol. 2, no. 4, pp. 140-148, 2009.
- [29] T. H. Markham, *Project Based Learning Handbook*, Buck Institute for Education, 2003.
- [30] Y. Gülbahar and H. Tinmaz, "Implementing Project-Based Learning And E-Portfolio Assessment In an Undergraduate Course," *Journal of Research on Technology in Education*, vol. 38, no. 3, pp. 309-327, Spring 2006.

# AUTOMATED RECOMMENDATION OF INFORMATION TO THE MEDIA BY THE IMPLEMENTATION OF WEB SEARCHING TECHNIQUE

Dr.Ashit kumar Dutta, Associate Professor, Shaqra University

## Abstract

Internet become the important media among the people all over the world. All other media depend on internet to gather information about the user navigational pattern and uses those information for their development. Web mining is the technology used for research carried out in internet. The notion of the research is to recommend the media to publish the frequently searched topics as news. The research uses google trends and hot trends data to find out the frequently searched topics by the user. An automated system is implemented to crawl items from the google trends and to recommend the same to the media.

**Keyword:** Internet, Recommendation system, feeds, web mining, Text mining

## I. INTRODUCTION

Web is a repository of information. People tends web to grow themselves by doing business. Many useful research gave a good shape to technology to utilise the web in a proper way for the development of society.[1][2] Web mining is the technology used to extract data from the web and provide knowledge from it.[3][4]

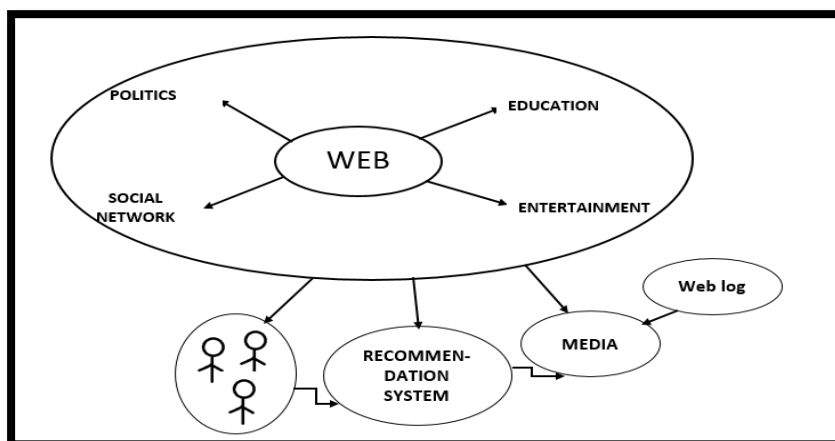


Figure 1 Web and Information recommendation system

## A. Google Trends

Google trends is one of the service offered by the google Inc. User can compare the volume of searches between two or three terms. Google allows user to track of various words and sentences searched through google. It has many features like organizing data in a pictorial format and customize the process according to the user needs. Google hot trends is the additional feature which give hourly analysis of the search made by the user all over the

world, user can filter the analysis by region and genre. Google trends is available as an application for the mobile users. User can install through apps store to analyse the trends of the user surfing the internet.[5]

### *B. Web Crawler*

Web crawler is the intelligent program to crawl the web content automatically from the web. It is also called as web spider.[6][7] Search engine uses robot to crawl and index websites in their database. Robot.txt are the file used by the website to be crawled by the web robots of search engine. Seeds are the list of URLs visited by the crawler.[8][9] These seeds are used to recursively crawl the websites according to the policies stored in the search engine or website robot.

### *C. Rich Site Summary (RSS) Feed*

Rss feeds are the short form of published information in the web. Millions of websites are work on the web and it is very difficult for the user to find the specific web page to get the desired information. Rss feeds are used to inter link websites and user can get the desired information through other websites. Xml format Rss feeds are available for the websites to publish it on their website.[10][11]

The objective of the research is to recommend the hot trends searches made by the user to the news media as well as display the same in the web portal..[12][13] A survey done by USA Today and the first amendment center found that 70% of people distrust the news media and said news media are biased. Sometimes important news were unnoticed by the new media and never brought to the society. .[13][14]The notion of the research is to find the average volume searches made by the internet users should be published in media..[15][16]

## II. REVIEW OF THE LITERATURE

Manish sethi and Geetika Gaur proposed a model to recommend news to the user. The work has analysed the different models of content based proposal and collaborative suggestion and made a cross over proposal frame work as an answer for the issues of news suggestion.[1]

Kathleen Tsoukalas et.al. have developed a system by implementing a fusion of web log mining technique that extracts and recommends frequently accessed terms to readers by utilizing information found in web query logs.[2]

Mariam Adedoyin – olowe et.al. have made a survey on data mining techniques for social network analysis. The research discussed different data mining techniques used in mining social networking sites data. The work has analysed the features of techniques used for social networking analysis.[3]

J. Liu et.al. proposed a news recommendation model based on click behaviour of the users. The research is based on the past click behaviour of the users. The system uses google news data to display according to the user profile. [4]

Abhinandan Das et.al. have proposed a method of collaborative filtering for generating generalized recommendations for users of google news. Their approach is easily adaptable for other applications with minimum modifications.[5]

D.Billsus and M. Pazzani developed a technique to create a model of the users preference from the user history in a classification learning form. They have trained the model using the user like and unlike data.[6]

Durairaj and muthukumar studied the different proposals and approaches that take users collective intelligence and navigation patterns. The research has compared various models used to recommend the news.[7]

Ujwala H. wanaskar et.al. proposed a method using weighted association rule mining algorithm and text mining. The method has produced good results comparing to the existing methods in the field.[8]

### III. METHODOLOGY

The research uses google trends data for the purpose of news recommendation. The data will be compared with the rss feeds extracted from the news media websites.[17][18][19] The data should be pre-processed and normalized before the comparison stage. The following figure 2 shows the framework of the research.

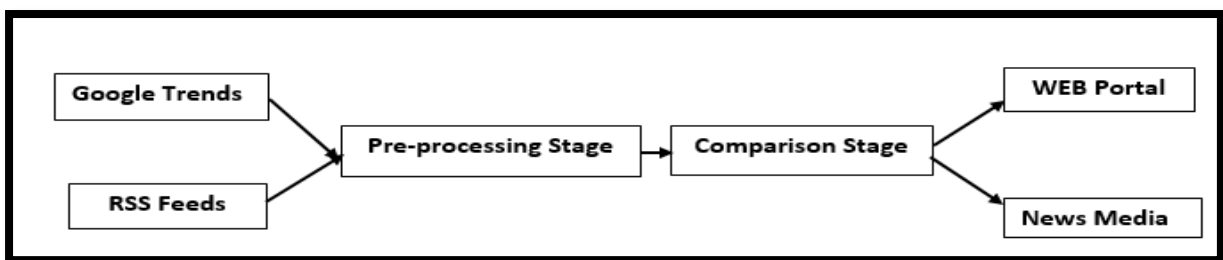


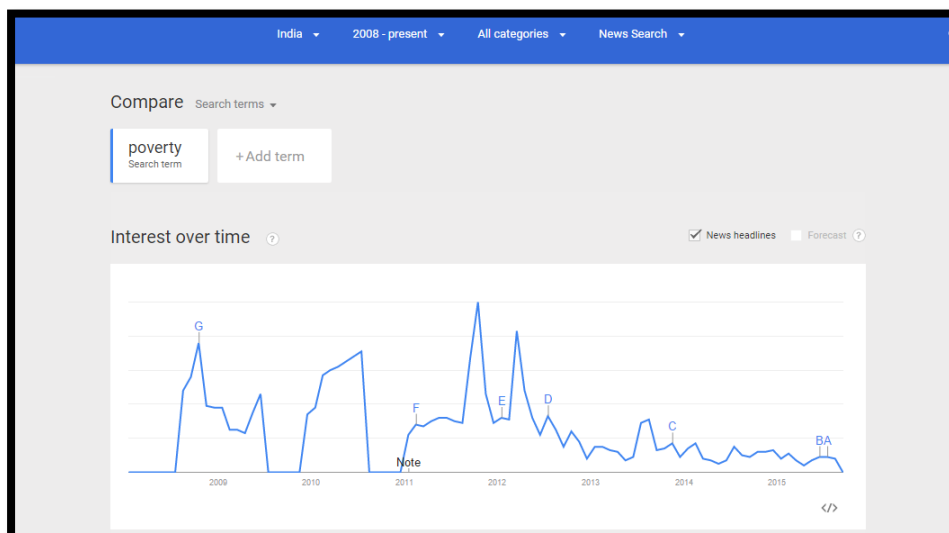
Figure 2 Framework of the Recommendation system

The comparison stage will generate the results and send as the recommendation to the news media and the web portal. Similar news and the data having good frequency will be discarded because that kind of news could be published in the news media. The research will search for the medium frequency data which is not similar with the rss feeds of the news media and publish it on web and send as the recommendation to the news media.

### IV. RESULTS AND DISCUSSION

The word “poverty” is keyed into the google trends topic search text box and for the period “2008 – present” was selected and news search is selected to retrieve the news related to the topic. The graph in the following figure

the  
of the  
topic



3 shows  
impact  
given  
for the  
selected  
period.

Figure 3 Google Trends graph for the keyword “poverty”

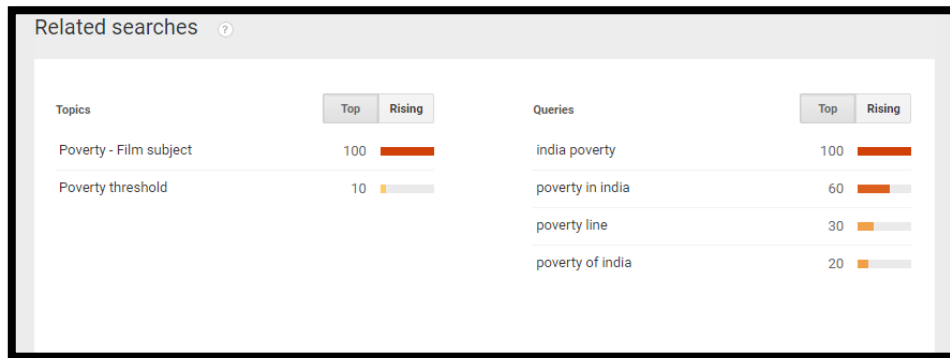


Figure 4 Related search topic for the keyword “poverty”

The volume of the search increased during the end of year 2011. The interesting point to be noted down in the context is the alphabets marked in the graph. The alphabets are news published in the new media. The highest volume of search occurred during the end of 2011 but that topic did not published by any of the news media in India. The lowest region in the graph is the present period shows two alphabets as the indication of news published in the news media. The figure 4 shows the related topic searched during the chosen period and these are the vital data could be published as a news in the media.

97	2015-08	8
98	2015-09	0
99		
100		
101	Top subregions for poverty	
102	Subregion	poverty
103	Delhi	100
104	Tamil Nadu	75
105	Karnataka	60
106	Uttar Pradesh	58
107	West Bengal	49
108	Maharashtra	48
109	Andaman and Nicobar Islands	0
110	Andhra Pradesh	0
111	Arunachal Pradesh	0

Figure 5. Data in the spreadsheet

The reason could be the impact of technology in the internet. The data for the work were downloaded from the google trends as a csv file and imported in the Microsoft excel spreadsheet for the pre-process work. The figure 5 shows the spreadsheet of the data extracted from the google trends. In the pre-process stage, the meaning less data were deleted and normalized for the comparison stage. The rss feed of the different websites were collected as a xml file and pre-processed then compared with the trends data. The following figure 4.3 is the one of the rss feed collected for the research work.



```
<?xml version="1.0" encoding="UTF-8" ?>
<rss version="2.0" ?>
  <channel ?>
    <title>Indian man dragged by driver in US</title>
    <link>http://zeenews.india.com/news/india/indian-man-dragged-by-driver-in-us_1667168.html</link>
    <description>An Indian attendant at a gas station in the US was injured when he was dragged by a car after the driver allegedly tried to steal cash from him.</description>
    <comment>mailto:inewsonline@gmail.com</comment>
    <pubDate>Fri, 04 September 2015, 11:46 GMT +0:30</pubDate>
    <author></author>
    <guid isPermaLink="false">http://zeenews.india.com/news/india/indian-man-dragged-by-driver-in-us_1667168.html</guid>
  </channel>
  <item ?>
    <title>GROP stalemate to end soon, Govt's draft agreement ready, claim sources</title>
    <link>http://zeenews.india.com/news/india/grop-stalemate-to-end-soon-govts-draft-agreement-ready-claim-source_1667167.html</link>
    <description>The ongoing deadlock in the implementation of the 'One Rank, One Pension' (GROP) may end soon as the government is likely to make an announcement in this regard within the next 48 hours.</description>
  </item>
</rss>
```

Figure 6. RSS feed of the Zee news media

The topic “poverty india” and “poor india” topics are generated as the output from the system. The reason for the output is the lowest curve in the year 2015 and the topics were very much dissimilar from the rss feeds. Finally the news “NBR Reporter Jason walls said India’s GDP growth rate above china” and recommended to the media and published in the web portal. The different keyword fetches different kinds of news and that can be easily recommended to the news media.

## V. CONCLUSION

News recommendation system is the useful method of publishing unnoticed news to the society. Web contains lot of data and google trends is the excellent tool to represent the frequent searches made by the users. The rss feeds are the short notes of the published news from the news media. The research successfully compared both google trends and rss feeds data and generated the topics searched frequently and unnoticed from the media.

## REFERENCES

- [1] Manish Sethi and Geetika Gaur, “Web mining techniques for extraction of news”, IJAETMAS, Vol. 2, Iss. 4, April 2015, pp.34 – 47.
- [2] Kathleen Tsoukalas, Bin Zhou, Jian Pei and DavorCubranic, “ Personalizing entity detection and recommendation with a fusion of web log mining techniques”, 12th International Conference on Extending Database Technology, Saint Petersburg, Russia, March 24-26, 2009, Proceedings.
- [3] Mariam Adedoyin – Olowe, Mohammed Medhat Gaber and Frederic Stahl, “ A survey of data mining techniques for social network analysis”, Journal of Data Mining & Digital Humanities, 2014 (June 24, 2014).
- [4] J. Liu, P. Dolan, and E. Pedersen, "Personalized news recommendation based on click behavior," in Proc. of the 15th Int. Conf. on IUI, 2010, pp. 31-40.
- [5] A. Das, M. Datar, A. Garg, and S. Rajaram, "Google news personalization: scalable online collaborative filtering," in Proc. of the 16th Int. Conf. on World Wide Web, 2007, pp. 271-280.
- [6] D. Billsus and M. Pazzani, "A hybrid user model for news story classification," in Proc. of the 7th Int. Conf. on User Modeling, 1999.
- [7] M. Durairaj and K.Muthukumar, “ News recommendation systems using web mining: A study”, International journal of Engineering Trends and Technology, Vol.12(6),2014, pp. 293 – 299.
- [8] Ujwala H. Wanaskar, SheetalR.Vij and Debajyotimukhopadhyay,” A hybrid web recommendation system based on the improved association rule mining algorithm”, Journal of software engineering and applications, 2013, 6, pp.396 – 404.
- [9] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," Trans. on Know. and Data Eng., vol. 17, pp. 734-749, 2005. (Pubitemid 40860454)
- [10] X. Su and T. Khoshgoftaar, "A survey of collaborative filtering techniques," Adv. in Artif. Intell., pp. 2-2, January 2009.
- [11] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "GroupLens: an open architecture for collaborative filtering of netnews," in Proc. of the Conf. on CSCW, 1994, pp. 175-186. P. Lops, M. Gemmis, and G. Semeraro, "Content-based recommender systems: State of the art and trends," in Recommender Systems Handbook, 2011, pp. 73-105.
- [12] J. Konstan, B. Miller, D. Maltz, J. Herlocker, L. Gordon, and J. Riedl, "GroupLens: applying collaborative filtering to usenet news," Commun. ACM, vol. 40, pp. 77-87, March 1997. (Pubitemid 127441134)
- [13] K. Lang, "Newsweeder: Learning to filter netnews," in Proc. of the 12th Int. Conf. on Machine Learning, 1995, pp. 331-339.
- [14] J. Ahn, P. Brusilovsky, J. Grady, and D. He, "Open user profiles for adaptive news systems: help or harm?" in Proc. of the 16th Int. Conf. on WWW, 2007, pp. 11-20.
- [15] R. Burke, "Hybrid recommender systems: Survey and experiments," User Modeling and User-Adapted Interaction, vol. 12, pp. 331-370, November 2002.
- [16] L. Li, W. Chu, J. Langford, and R. Schapire, "A contextual-bandit approach to personalized news article recommendation," in Proc. of the 19th Int. Conf. on World Wide Web, 2010, pp. 661-670.
- [17] G. Shani, D. Heckerman, and R. Brafman, "An mdp-based recommender system," J. Mach. Learn. Res., vol. 6, pp. 1265-1295, 2005.
- [18] T. Hofmann, "Probabilistic latent semantic indexing," in Proc. of the 22nd Int. Conf. on Research and Development in Info. Retrieval, 1999, pp. 50-57.
- [19] D. Blei, A. Ng, and M. Jordan, "Latent dirichlet allocation," J. MLR, vol. 3, pp. 993-1022, 2003.

# Comparison of Euclidean Distance Function and Manhattan Distance Function Using K-Medoids

Md. Mohibullah  
Student (M.Sc. - Thesis)  
Department of Computer Science  
& Engineering  
Comilla University  
Comilla, Bangladesh

Md. Zakir Hossain  
Assistant Professor  
Department of Computer Science  
& Engineering  
Comilla University  
Comilla, Bangladesh

Mahmudul Hasan  
Assistant Professor  
Department of Computer Science  
& Engineering  
Comilla University  
Comilla, Bangladesh.

**Abstract--Clustering is one kind of unsupervised learning methods. K-medoids is one of the partitioning clustering algorithms and it is also a distance based clustering. Distance measure is an important component of a clustering algorithm to measure the distances between data points. In this thesis paper, a comparison between Euclidean distance function and Manhattan distance function by using K-medoids has been made. To make this comparison, an instance of seven objects of a data set has been taken. Finally, we will show the simulation results in the result section of this paper.**

**Keywords--** Clustering, K-medoids, Manhattan distance function, Euclidean distance function.

## I. INTRODUCTION

Unsupervised learning works on a given set of records (e.g. observations or variables) with no attribute and organize them into groups, without advance knowledge of the definitions of the groups [1]. Clustering is one of the most important unsupervised learning techniques. Clustering, also known as cluster analysis), aims to organize a collection of data items into clusters, such that items within a cluster are more “similar” to each other than they are to items in the other clusters [2]. Clustering methods can be divided into two basic types: hierarchical and partition clustering [3]. There are many partition-based algorithms such as K-Means, K-Medoids and Fuzzy C-Means clustering etc.

The k-means method uses centroid to represent the cluster and it is sensitive to outliers. This means, a data object with an extremely large value may disrupt the distribution of data. K-medoids method overcomes this problem by using medoids to represent the cluster rather than centroid. A medoid is the most centrally located data object in a cluster [4].

## II. THE REASON BEHIND CHOOSING K-MEDIOIDS ALGORITHM

### 1. K-medoid is more flexible

First of all, k-medoids can be used with any similarity measure. K-means however, may fail to converge - it really must only be used with distances that are consistent with the mean. So e.g. Absolute Pearson Correlation must not be used with k-means, but it works well with k-medoids.

### 2. Robustness of medoid

Secondly, the medoid as used by k-medoids is roughly comparable to the median. It is a more robust estimate of a representative point than the mean as used in k-means.

## III. K-MEDIOIDS ALGORITHM (PAM-PARTITIONING AROUND MEDIOIDS)

Algorithm [4, 6]

Input

K: the number of clusters

D: a data set containing n objects

Output: A set of k clusters.

Method

1. Compute distance (cost) so as to associate each data point to its nearest medoid using Manhattan distance and/or Euclidean distance.
2. for each medoid  $m$ 
  1. for each non-medoid data point  $o$ 
    1. Swap  $m$  and  $o$  and compute the total cost of the configuration
3. Select the configuration with the lowest cost.
4. Repeat steps 1 to 3 until there is no change in the medoid.



#### IV. DEMONSTRATION OF K-MEDOIDS

We will see the clustering of data set with an example for k-medoid algorithm using both the Manhattan distance and Euclidean distance.

For Instance: Consider a data set of seven objects as follows:

Serial No	Variable-1	Variable-2
1 (X <sub>1</sub> )	1.0	1.0
2 (X <sub>2</sub> )	1.5	2.0
3 (X <sub>3</sub> )	3.0	4.0
4 (X <sub>4</sub> )	5.0	7.0
5 (X <sub>5</sub> )	3.5	5.0
6 (X <sub>6</sub> )	4.5	5.0
7 (X <sub>7</sub> )	3.5	4.5

Table 1: A data set of seven objects

The following shows the scatter diagram of the above data set.

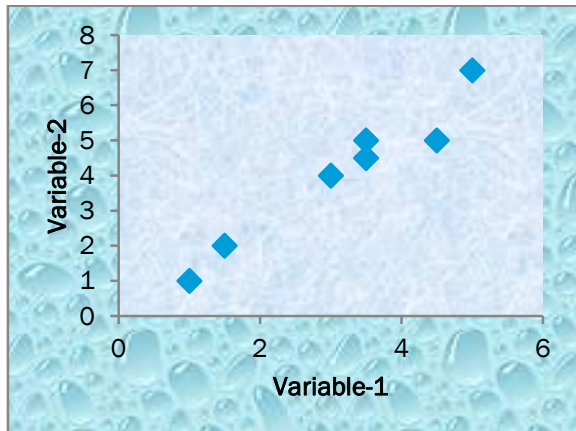


Figure 1: Distribution of data

#### V. USING MANHATTAN DISTANCE

##### Step 1

Consider the number of clusters is two i.e., k=2 and initialize k centers.

Let us assume  $c_1 = (1, 1)$  and  $c_2 = (5, 7)$

So here  $x_1$  and  $x_4$  are selected as medoids.

Calculate distance so as to associate each data object to its nearest medoid. Cost is calculated using Manhattan Distance. Costs to the nearest medoid are shown bold in the table.

i	C <sub>1</sub>		Data objects (X <sub>i</sub> )		Cost (distance)
2	1.0	1.0	1.5	2.0	$ 1.0 - 1.5  +  1.0 - 2.0  = \mathbf{1.5}$
3	1.0	1.0	3.0	4.0	$ 1.0 - 3.0  +  1.0 - 4.0  = \mathbf{5.0}$
5	1.0	1.0	3.5	5.0	$ 1.0 - 3.5  +  1.0 - 5.0  = 6.5$
6	1.0	1.0	4.5	5.0	$ 1.0 - 4.5  +  1.0 - 5.0  = 7.5$
7	1.0	1.0	3.5	4.5	$ 1.0 - 3.5  +  1.0 - 4.5  = 6.0$

i	C <sub>2</sub>		Data objects (X <sub>i</sub> )		Cost (distance)
2	5.0	7.0	1.5	2.0	$ 5.0 - 1.5  +  7.0 - 2.0  = 8.5$
3	5.0	7.0	3.0	4.0	$ 5.0 - 3.0  +  7.0 - 4.0  = \mathbf{5.0}$
5	5.0	7.0	3.5	5.0	$ 5.0 - 3.5  +  7.0 - 5.0  = \mathbf{3.5}$
6	5.0	7.0	4.5	5.0	$ 5.0 - 4.5  +  7.0 - 5.0  = \mathbf{2.5}$
7	5.0	7.0	3.5	4.5	$ 5.0 - 3.5  +  7.0 - 4.5  = \mathbf{4.0}$

Since the cost for X<sub>2</sub> is not changed. So we can keep it in cluster-1. Then the clusters become:

Cluster-1 =  $\{(1, 1), (1.5, 2), (3, 4)\}$  i.e.  $\{X_1, X_2, X_3\}$

Cluster-2 =  $\{(5, 7), (3.5, 5), (4.5, 5), (3.5, 4.5)\}$  i.e.  $\{X_4, X_5, X_6, X_7\}$

Since the points (3.5, 5), (4.5, 5), and (3.5, 4.5) are closer to C<sub>2</sub>, hence they form one cluster and the remaining points form another cluster C<sub>1</sub>.

So the total cost involved is 16.5.

Where the cost is calculated by following formula:

$$\text{Cost}(X, C) = \sum_{i=0}^n |X_i - C_i|$$

Where  $x$  is any data object,  $c$  is the medoid, and  $n$  is the dimension of the object which in this case is 2.

Total cost is the summation of the minimum cost of data object from its medoid in its cluster so here:

$$\text{Total cost} = (1.5 + 5) + (3.5 + 2.5 + 4) = 16.5$$

*Step 2*

Select one of the nonmedoids  $O'$ . Let us assume  $O' = (4.5, 5.0)$ . So now the medoids are  $C_1 (1, 1)$  and  $O' (4.5, 5)$ .

Again, calculate distance so as to associate each data object to its nearest medoid. Costs to the nearest medoid are shown bold in the table.

i	$C_1$		Data objects ( $X_i$ )		Cost (distance)
2	1.0	1.0	1.5	2.0	$ 1.0 - 1.5  +  1.0 - 2.0  = \mathbf{1.5}$
3	1.0	1.0	3.0	4.0	$ 1.0 - 3.0  +  1.0 - 4.0  = 5.0$
4	1.0	1.0	5.0	7.0	$ 1.0 - 5.0  +  1.0 - 7.0  = 10$
5	1.0	1.0	3.5	5.0	$ 1.0 - 3.5  +  1.0 - 5.0  = 6.5$
7	1.0	1.0	3.5	4.5	$ 1.0 - 3.5  +  1.0 - 4.5  = 6.0$

i	$O'$		Data objects ( $X_i$ )		Cost (distance)
2	4.5	5.0	1.5	2.0	$ 4.5 - 1.5  +  5.0 - 2.0  = 6$
3	4.5	5.0	3.0	4.0	$ 4.5 - 3.0  +  5.0 - 4.0  = \mathbf{2.5}$
4	4.5	5.0	5.0	7.0	$ 4.5 - 5.0  +  5.0 - 7.0  = \mathbf{2.5}$
5	4.5	5.0	3.5	5.0	$ 4.5 - 3.5  +  5.0 - 5.0  = \mathbf{1}$
7	4.5	5.0	3.5	4.5	$ 4.5 - 3.5  +  5.0 - 4.5  = \mathbf{1.5}$

From the step 2, we get the following clusters:

Cluster-1 =  $\{(1, 1), (1.5, 2)\}$  i.e.  $\{X_1, X_2\}$

Cluster-2 =  $\{(3, 4), (5, 7), (3.5, 5), (4.5, 5), (3.5, 4.5)\}$   
i.e.  $\{X_3, X_4, X_5, X_6, X_7\}$

The total cost =  $1.5 + 2.5 + 2.5 + 1 + 1.5 = 9$

*Cost comparison*

From step 1 and step 2, we get the total cost are 16.5 and 9 respectively. So cost of swapping medoid from  $C_2$  to  $O'$  is

$S = \text{Current total cost} - \text{Past total cost}$

$$= 9 - 16.5$$

$$= -7.5 < 0$$

So moving would be a good idea and the previous choice was a bad idea. Now we will try to again to certain for the clustering.

*Step 3*

Select another nonmedoid  $P'$ . Let us assume  $P' = (3.5, 4.5)$ . So now the medoids are  $C_1 (1, 1)$  and  $P' (3.5, 4.5)$ .

Again, calculate distance so as to associate each data object to its nearest medoid. Cost is calculated using Manhattan Distance. Costs to the nearest medoid are shown bold in the table.

i	$C_1$		Data objects ( $X_i$ )		Cost (distance)
2	1.0	1.0	1.5	2.0	$ 1.0 - 1.5  +  1.0 - 2.0  = \mathbf{1.5}$
3	1.0	1.0	3.0	4.0	$ 1.0 - 3.0  +  1.0 - 4.0  = 5.0$
4	1.0	1.0	5.0	7.0	$ 1.0 - 5.0  +  1.0 - 7.0  = 10$
5	1.0	1.0	3.5	5.0	$ 1.0 - 3.5  +  1.0 - 5.0  = 6.5$
6	1.0	1.0	4.5	5.0	$ 1.0 - 4.5  +  1.0 - 5.0  = 7.5$

i	$P'$		Data objects ( $X_i$ )		Cost (distance)
2	3.5	4.5	1.5	2.0	$ 3.5 - 1.5  +  4.5 - 2.0  = 4.5$
3	3.5	4.5	3.0	4.0	$ 3.5 - 3.0  +  4.5 - 4.0  = \mathbf{1.0}$
4	3.5	4.5	5.0	7.0	$ 3.5 - 5.0  +  4.5 - 7.0  = \mathbf{4.0}$
5	3.5	4.5	3.5	5.0	$ 3.5 - 3.5  +  4.5 - 5.0  = \mathbf{0.5}$
6	3.5	4.5	4.5	5.0	$ 3.5 - 4.5  +  4.5 - 5.0  = \mathbf{1.5}$

From the step 3, we get the following clusters:

Cluster-1 =  $\{(1, 1), (1.5, 2)\}$  i.e.  $\{X_1, X_2\}$

Cluster-2 =  $\{(3, 4), (5, 7), (3.5, 5), (4.5, 5), (3.5, 4.5)\}$   
i.e.  $\{X_3, X_4, X_5, X_6, X_7\}$

The total cost =  $1.5 + 1.0 + 4.0 + 0.5 + 1.5 = 8.5$

*Cost comparison*

From step 2 and step 3, we get the total cost are 9.0 and 8.5 respectively. So cost of swapping medoid from  $O'$  to  $P'$  is

$S = \text{Current total cost} - \text{Past total cost}$

$$= 8.5 - 9.0$$

$$= -0.5 < 0$$

So moving would be a good idea.

*Step 4*

Select another nonmedoid  $Q'$ . Let us assume  $Q' = (3.5, 5.0)$ . So now the medoids are  $C_1 (1, 1)$  and  $Q' (3.5, 5.0)$ .

Again, calculate distance so as to associate each data object to its nearest medoid. Costs to the nearest medoid are shown bold in the table.

i	$C_1$		Data objects ( $X_i$ )		Cost (distance)
2	1.0	1.0	1.5	2.0	$ 1.0 - 1.5  +  1.0 - 2.0  = \mathbf{1.5}$
3	1.0	1.0	3.0	4.0	$ 1.0 - 3.0  +  1.0 - 4.0  = 5.0$
4	1.0	1.0	5.0	7.0	$ 1.0 - 5.0  +  1.0 - 7.0  = 10$
6	1.0	1.0	4.5	5.0	$ 1.0 - 4.5  +  1.0 - 5.0  = 7.5$
7	1.0	1.0	3.5	4.5	$ 1.0 - 3.5  +  1.0 - 4.5  = 6.0$

i	$Q'$		Data objects ( $X_i$ )		Cost (distance)
2	3.5	5.0	1.5	2.0	$ 3.5 - 1.5  +  5.0 - 2.0  = 5.0$
3	3.5	5.0	3.0	4.0	$ 3.5 - 3.0  +  5.0 - 4.0  = \mathbf{1.5}$
4	3.5	5.0	5.0	7.0	$ 3.5 - 5.0  +  5.0 - 7.0  = \mathbf{3.5}$
6	3.5	5.0	4.5	5.0	$ 3.5 - 4.5  +  5.0 - 5.0  = \mathbf{1.0}$
7	3.5	5.0	3.5	4.5	$ 3.5 - 3.5  +  5.0 - 4.5  = \mathbf{0.5}$

From the step 4, we get the following clusters:

Cluster-1 =  $\{(1, 1), (1.5, 2)\}$  i.e.  $\{X_1, X_2\}$

Cluster-2 =  $\{(3, 4), (5, 7), (3.5, 5), (4.5, 5), (3.5, 4.5)\}$   
i.e.  $\{X_3, X_4, X_5, X_6, X_7\}$

The total cost =  $1.5 + 1.5 + 3.5 + 1.0 + 0.5 = 8.0$

*Cost comparison*

From step 3 and step 4, we get the total cost are 8.5 and 8.0 respectively. So cost of swapping medoid from  $P'$  to  $Q'$  is

$$S = \text{Current total cost} - \text{Past total cost}$$

$$= 8.0 - 8.5$$

$$= -0.5 < 0$$

So moving would be a good idea.

*Step 5*

Select another nonmedoid  $R'$ . Let us assume  $R' = (3.0, 4.0)$ . So now the medoids are  $C_1 (1, 1)$  and  $R' (3.0, 4.0)$ .

Again, calculate distance so as to associate each data object to its nearest medoid. Costs to the nearest medoid are shown bold in the table.

i	$C_1$		Data objects ( $X_i$ )		Cost (distance)
2	1.0	1.0	1.5	2.0	$ 1.0 - 1.5  +  1.0 - 2.0  = \mathbf{1.5}$
4	1.0	1.0	5.0	7.0	$ 1.0 - 5.0  +  1.0 - 7.0  = 10$
5	1.0	1.0	3.5	5.0	$ 1.0 - 3.5  +  1.0 - 5.0  = 6.5$
6	1.0	1.0	4.5	5.0	$ 1.0 - 4.5  +  1.0 - 5.0  = 7.5$
7	1.0	1.0	3.5	4.5	$ 1.0 - 3.5  +  1.0 - 4.5  = 6.0$

i	$R'$		Data objects ( $X_i$ )		Cost (distance)
2	3.0	4.0	1.5	2.0	$ 3.0 - 1.5  +  4.0 - 2.0  = 3.5$
4	3.0	4.0	5.0	7.0	$ 3.0 - 5.0  +  4.0 - 7.0  = \mathbf{5.0}$
5	3.0	4.0	3.5	5.0	$ 3.0 - 3.5  +  4.0 - 5.0  = \mathbf{1.5}$
6	3.0	4.0	4.5	5.0	$ 3.0 - 4.5  +  4.0 - 5.0  = \mathbf{2.5}$
7	3.0	4.0	3.5	4.5	$ 3.0 - 3.5  +  4.0 - 4.5  = \mathbf{1.0}$

From the step 5, we get the following clusters:

Cluster-1 =  $\{(1, 1), (1.5, 2)\}$  i.e.  $\{X_1, X_2\}$

Cluster-2 =  $\{(3, 4), (5, 7), (3.5, 5), (4.5, 5), (3.5, 4.5)\}$   
i.e.  $\{X_3, X_4, X_5, X_6, X_7\}$

The total cost =  $1.5 + 5.0 + 1.5 + 2.5 + 1.0 = 11.5$

*Cost comparison*

From step 4 and step 5 we get the total cost are 8.0 and 11.5 respectively. So cost of swapping medoid from  $Q'$  to  $R'$  is

$$S = \text{Current total cost} - \text{Past total cost}$$

$$= 11.5 - 8.0$$

$$= 3.5 > 0$$

So moving would be a bad idea and the previous choice was a good idea.

*Step 6*

Select another nonmedoid  $S'$ . Let us assume  $S' = (1.5, 2.0)$ . So now the medoids are  $C_1(1, 1)$  and  $S'(1.5, 2.0)$ .

Again, calculate distance so as to associate each data object to its nearest medoid. Costs to the nearest medoid are shown bold in the table.

i	$C_1$		Data objects ( $X_i$ )		Cost (distance)
3	1.0	1.0	3.0	4.0	$ 1.0 - 3.0  +  1.0 - 4.0  = 5.0$
4	1.0	1.0	5.0	7.0	$ 1.0 - 5.0  +  1.0 - 7.0  = 10$
5	1.0	1.0	3.5	5.0	$ 1.0 - 3.5  +  1.0 - 5.0  = 6.5$
6	1.0	1.0	4.5	5.0	$ 1.0 - 4.5  +  1.0 - 5.0  = 7.5$
7	1.0	1.0	3.5	4.5	$ 1.0 - 3.5  +  1.0 - 4.5  = 6.0$

i	$S'$		Data objects ( $X_i$ )		Cost (distance)
3	1.5	2.0	3.0	4.0	$ 1.5 - 3.0  +  2.0 - 4.0  = \mathbf{3.5}$
4	1.5	2.0	5.0	7.0	$ 1.5 - 5.0  +  2.0 - 7.0  = \mathbf{8.5}$
5	1.5	2.0	3.5	5.0	$ 1.5 - 3.5  +  2.0 - 5.0  = \mathbf{4.5}$
6	1.5	2.0	4.5	5.0	$ 1.5 - 4.5  +  2.0 - 5.0  = \mathbf{6.0}$
7	1.5	2.0	3.5	4.5	$ 1.5 - 3.5  +  2.0 - 4.5  = \mathbf{4.0}$

From the step 6, we get the following clusters:

Cluster-1 =  $\{(1, 1)\}$  i.e.  $\{X_1\}$

Cluster-2 =  $\{(1.5, 2), (3, 4), (5, 7), (3.5, 5), (4.5, 5), (3.5, 4.5)\}$  i.e.  $\{X_2, X_3, X_4, X_5, X_6, X_7\}$

The total cost =  $3.5 + 8.5 + 4.5 + 6.0 + 4.0 = 26.5$

*Cost comparison*

From step 4 and step 6 we get the total cost are 8.0 and 26.5 respectively. So cost of swapping medoid from  $Q'$  to  $S'$  is

$S = \text{Current total cost} - \text{Past total cost}$

$$= 26.5 - 8.0$$

$$= 18.5 > 0$$

So moving would be a bad idea and the choice in step 4 was a good idea. So the configuration does not change after step 4 and algorithm terminates here (i.e. there is no change in the medoids- the medoids are  $X_1$  and  $X_5$ ).

VI. USING EUCLIDEAN DISTANCE

*Step 1*

Consider the number of clusters is two i.e.,  $k=2$  and initialize  $k$  centers.

Let us assume  $c_1 = (1, 1)$  and  $c_2 = (5, 7)$

So here  $x_1$  and  $x_4$  are selected as medoids.

Calculate distance so as to associate each data object to its nearest medoid. Cost is calculated using Euclidean Distance. Costs to the nearest medoid are shown bold in the table.

i	$C_1$		Data objects ( $X_i$ )		Cost (distance)
2	1.0	1.0	1.5	2.0	$\sqrt{ 1.0 - 1.5 ^2 +  1.0 - 2.0 ^2} = \mathbf{1.118}$
3	1.0	1.0	3.0	4.0	$\sqrt{ 1.0 - 3.0 ^2 +  1.0 - 4.0 ^2} = \mathbf{3.606}$
5	1.0	1.0	3.5	5.0	$\sqrt{ 1.0 - 3.5 ^2 +  1.0 - 5.0 ^2} = 4.717$
6	1.0	1.0	4.5	5.0	$\sqrt{ 1.0 - 4.5 ^2 +  1.0 - 5.0 ^2} = 5.315$
7	1.0	1.0	3.5	4.5	$\sqrt{ 1.0 - 3.5 ^2 +  1.0 - 4.5 ^2} = 4.301$

i	$C_2$		Data objects ( $X_i$ )		Cost (distance)
2	5.0	7.0	1.5	2.0	$\sqrt{ 5.0 - 1.5 ^2 +  7.0 - 2.0 ^2} = 6.103$
3	5.0	7.0	3.0	4.0	$\sqrt{ 5.0 - 3.0 ^2 +  7.0 - 4.0 ^2} = \mathbf{3.606}$
5	5.0	7.0	3.5	5.0	$\sqrt{ 5.0 - 3.5 ^2 +  7.0 - 5.0 ^2} = \mathbf{2.5}$

6	5.0	7.0	4.5	5.0	$\sqrt{ 5.0 - 4.5 ^2 +  7.0 - 5.0 ^2}$ <b>= 2.062</b>
7	5.0	7.0	3.5	4.5	$\sqrt{ 5.0 - 3.5 ^2 +  7.0 - 4.5 ^2}$ <b>= 2.915</b>

Since the cost for  $X_2$  is not changed. So we can keep it in cluster-1. Then the clusters become:

Cluster-1=  $\{(1, 1), (1.5, 2), (3, 4)\}$  i.e.  $\{X_1, X_2, X_3\}$

Cluster-2 =  $\{(5, 7), (3.5, 5), (4.5, 5), (3.5, 4.5)\}$  i.e.  $\{X_4, X_5, X_6, X_7\}$

Since the points (3.5, 5), (4.5, 5), and (3.5, 4.5) are closer to  $C_2$ , hence they form one cluster and the remaining points form another cluster  $C_1$ .

Total cost is the summation of the minimum cost of data object from its medoid in its cluster so here:

$$\text{Total cost} = (1.118 + 3.606) + (2.5 + 2.062 + 2.915) = 12.201$$

### Step 2

Select one of the nonmedoids  $O'$ . Let us assume  $O' = (4.5, 5.0)$ . So now the medoids are  $C_1 (1, 1)$  and  $O' (4.5, 5)$ .

Again, calculate distance so as to associate each data object to its nearest medoid. Costs to the nearest medoid are shown bold in the table.

i	$C_1$		Data objects ( $X_i$ )		Cost (distance)
2	1.0	1.0	1.5	2.0	$\sqrt{ 1.0 - 1.5 ^2 +  1.0 - 2.0 ^2}$ <b>= 1.118</b>
3	1.0	1.0	3.0	4.0	$\sqrt{ 1.0 - 3.0 ^2 +  1.0 - 4.0 ^2}$ <b>= 3.606</b>
4	1.0	1.0	5.0	7.0	$\sqrt{ 1.0 - 5.0 ^2 +  1.0 - 7.0 ^2}$ <b>= 7.211</b>
5	1.0	1.0	3.5	5.0	$\sqrt{ 1.0 - 3.5 ^2 +  1.0 - 5.0 ^2}$ <b>= 4.717</b>
7	1.0	1.0	3.5	4.5	$\sqrt{ 1.0 - 3.5 ^2 +  1.0 - 4.5 ^2}$ <b>= 4.301</b>

i	$O'$	Data objects ( $X_i$ )	Cost (distance)

2	4.5	5.0	1.5	2.0	$\sqrt{ 4.5 - 1.5 ^2 +  5.0 - 2.0 ^2}$ <b>= 4.243</b>
3	4.5	5.0	3.0	4.0	$\sqrt{ 4.5 - 3.0 ^2 +  5.0 - 4.0 ^2}$ <b>= 1.803</b>
4	1.0	1.0	5.0	7.0	$\sqrt{ 4.5 - 5.0 ^2 +  5.0 - 7.0 ^2}$ <b>= 2.062</b>
5	4.5	5.0	3.5	5.0	$\sqrt{ 4.5 - 3.5 ^2 +  5.0 - 5.0 ^2}$ <b>= 1.0</b>
7	4.5	5.0	3.5	4.5	$\sqrt{ 4.5 - 3.5 ^2 +  5.0 - 4.5 ^2}$ <b>= 1.118</b>

From the step 2, we get the following clusters:

Cluster-1=  $\{(1, 1), (1.5, 2)\}$  i.e.  $\{X_1, X_2\}$

Cluster-2 =  $\{(3, 4), (5, 7), (3.5, 5), (4.5, 5), (3.5, 4.5)\}$  i.e.  $\{X_3, X_4, X_5, X_6, X_7\}$

$$\text{The total cost} = 1.118 + 1.803 + 2.062 + 1.0 + 1.118 = 7.101$$

### Cost comparison

From step 1 and step 2, we get the total cost are 12.201 and 7.101 respectively. So cost of swapping medoid from  $C_2$  to  $O'$  is

$$S = \text{Current total cost} - \text{Past total cost}$$

$$= 7.101 - 12.201$$

$$= -5.1 < 0$$

So moving would be a good idea and the previous choice was a bad idea. Now we will try to again to certain for the clustering.

### Step 3

Select another nonmedoid  $P'$ . Let us assume  $P' = (3.5, 4.5)$ . So now the medoids are  $C_1 (1, 1)$  and  $P' (3.5, 4.5)$ .

Again, calculate distance so as to associate each data object to its nearest medoid. Costs to the nearest medoid are shown bold in the table.

i	$C_1$	Data objects ( $X_i$ )	Cost (distance)

2	1.0	1.0	1.5	2.0	$\sqrt{ 1.0 - 1.5 ^2 +  1.0 - 2.0 ^2}$ = <b>1.118</b>
3	1.0	1.0	3.0	4.0	$\sqrt{ 1.0 - 3.0 ^2 +  1.0 - 4.0 ^2}$ = 3.606
4	1.0	1.0	5.0	7.0	$\sqrt{ 1.0 - 5.0 ^2 +  1.0 - 7.0 ^2}$ = 7.211
5	1.0	1.0	3.5	5.0	$\sqrt{ 1.0 - 3.5 ^2 +  1.0 - 5.0 ^2}$ = 4.717
6	1.0	1.0	4.5	5.0	$\sqrt{ 1.0 - 4.5 ^2 +  1.0 - 5.0 ^2}$ = 5.315

i	P'		Data objects (X <sub>i</sub> )		Cost (distance)
2	3.5	4.5	1.5	2.0	$\sqrt{ 3.5 - 1.5 ^2 +  4.5 - 2.0 ^2}$ = 3.202
3	3.5	4.5	3.0	4.0	$\sqrt{ 3.5 - 3.0 ^2 +  4.5 - 4.0 ^2}$ = <b>0.707</b>
4	3.5	4.5	5.0	7.0	$\sqrt{ 3.5 - 5.0 ^2 +  4.5 - 7.0 ^2}$ = <b>2.915</b>
5	3.5	4.5	3.5	5.0	$\sqrt{ 3.5 - 3.5 ^2 +  4.5 - 5.0 ^2}$ = <b>0.5</b>
6	3.5	4.5	4.5	5.0	$\sqrt{ 3.5 - 4.5 ^2 +  4.5 - 5.0 ^2}$ = <b>1.118</b>

From the step 3, we get the following clusters:

Cluster-1= {(1, 1), (1.5, 2)} i.e. {X<sub>1</sub>, X<sub>2</sub>}

Cluster-2 = {(3, 4), (5, 7), (3.5, 5), (4.5, 5), (3.5, 4.5)}  
i.e. {X<sub>3</sub>, X<sub>4</sub>, X<sub>5</sub>, X<sub>6</sub>, X<sub>7</sub>}

The total cost= 1.118 + 0.707 + 2.915 + 0.5 + 1.118 = 6.358

#### Cost comparison

From step 2 and step 3, we get the total cost are 7.101 and 6.358 respectively. So cost of swapping medoid from O' to P' is

S= Current total cost – Past total cost

$$= 6.358 - 7.101$$

$$= -0.743 < 0$$

So moving would be a good idea.

#### Step 4

Select another nonmedoid Q'. Let us assume Q' = (3.5, 5). So now the medoids are C<sub>1</sub> (1, 1) and Q' (3.5, 5).

Again, calculate distance so as to associate each data object to its nearest medoid. Costs to the nearest medoid are shown bold in the table.

i	C <sub>1</sub>		Data objects (X <sub>i</sub> )		Cost (distance)
2	1.0	1.0	1.5	2.0	$\sqrt{ 1.0 - 1.5 ^2 +  1.0 - 2.0 ^2}$ = <b>1.118</b>
3	1.0	1.0	3.0	4.0	$\sqrt{ 1.0 - 3.0 ^2 +  1.0 - 4.0 ^2}$ = 3.606
4	1.0	1.0	5.0	7.0	$\sqrt{ 1.0 - 5.0 ^2 +  1.0 - 7.0 ^2}$ = 7.211
6	1.0	1.0	4.5	5.0	$\sqrt{ 1.0 - 4.5 ^2 +  1.0 - 5.0 ^2}$ = 5.315
7	1.0	1.0	3.5	4.5	$\sqrt{ 1.0 - 3.5 ^2 +  1.0 - 4.5 ^2}$ = 4.301

i	Q'		Data objects (X <sub>i</sub> )		Cost (distance)
2	3.5	5.0	1.5	2.0	$\sqrt{ 3.5 - 1.5 ^2 +  5.0 - 2.0 ^2}$ = 3.606
3	3.5	5.0	3.0	4.0	$\sqrt{ 3.5 - 3.0 ^2 +  5.0 - 4.0 ^2}$ = <b>1.118</b>
4	3.5	5.0	5.0	7.0	$\sqrt{ 3.5 - 5.0 ^2 +  5.0 - 7.0 ^2}$ = <b>2.50</b>
6	3.5	5.0	4.5	5.0	$\sqrt{ 3.5 - 4.5 ^2 +  5.0 - 5.0 ^2}$ = <b>1.00</b>
7	3.5	5.0	3.5	4.5	$\sqrt{ 3.5 - 3.5 ^2 +  5.0 - 4.5 ^2}$ = <b>0.50</b>

From the step 4, we get the following clusters:

Cluster-1= {(1, 1), (1.5, 2)} i.e. {X<sub>1</sub>, X<sub>2</sub>}

Cluster-2 = {(3, 4), (5, 7), (3.5, 5), (4.5, 5), (3.5, 4.5)}  
i.e. {X<sub>3</sub>, X<sub>4</sub>, X<sub>5</sub>, X<sub>6</sub>, X<sub>7</sub>}

The total cost= 1.118 + 1.118 + 2.5 + 1.0 + 0.5 = 6.236

#### Cost comparison

From step 3 and step 4, we get the total cost are 6.358 and 6.236 respectively. So cost of swapping medoid from P' to Q' is

S= Current total cost – Past total cost

$$= 6.236 - 6.358$$

$$= -0.122 < 0$$

So moving would be a good idea.

Step 5

Select another nonmedoid R'. Let us assume R' = (3.0, 4.0). So now the medoids are C<sub>1</sub> (1, 1) and R' (3.0, 4.0).

Again, calculate distance so as to associate each data object to its nearest medoid. Costs to the nearest medoid are shown bold in the table.

i	C <sub>1</sub>		Data objects (X <sub>i</sub> )		Cost (distance)
2	1.0	1.0	1.5	2.0	$\sqrt{ 1.0 - 1.5 ^2 +  1.0 - 2.0 ^2}$ <b>= 1.118</b>
4	1.0	1.0	5.0	7.0	$\sqrt{ 1.0 - 5.0 ^2 +  1.0 - 7.0 ^2}$ <b>= 7.211</b>
5	1.0	1.0	3.5	5.0	$\sqrt{ 1.0 - 3.5 ^2 +  1.0 - 5.0 ^2}$ <b>= 4.717</b>
6	1.0	1.0	4.5	5.0	$\sqrt{ 1.0 - 4.5 ^2 +  1.0 - 5.0 ^2}$ <b>= 5.315</b>
7	1.0	1.0	3.5	4.5	$\sqrt{ 1.0 - 3.5 ^2 +  1.0 - 4.5 ^2}$ <b>= 4.301</b>

i	R'		Data objects (X <sub>i</sub> )		Cost (distance)
2	3.0	4.0	1.5	2.0	$\sqrt{ 3.0 - 1.5 ^2 +  4.0 - 2.0 ^2}$ <b>= 2.5</b>
4	3.0	4.0	5.0	7.0	$\sqrt{ 3.0 - 5.0 ^2 +  4.0 - 7.0 ^2}$ <b>= 3.606</b>
5	3.0	4.0	3.5	5.0	$\sqrt{ 3.0 - 3.5 ^2 +  4.0 - 5.0 ^2}$ <b>= 1.118</b>
6	3.0	4.0	4.5	5.0	$\sqrt{ 3.0 - 4.5 ^2 +  4.0 - 5.0 ^2}$ <b>= 1.803</b>
7	3.0	4.0	3.5	4.5	$\sqrt{ 3.0 - 3.5 ^2 +  4.0 - 4.5 ^2}$ <b>= 0.707</b>

From the step 5, we get the following clusters:

Cluster-1= {(1, 1), (1.5, 2)} i.e. {X<sub>1</sub>, X<sub>2</sub>}

Cluster-2 = {(3, 4), (5, 7), (3.5, 5), (4.5, 5), (3.5, 4.5)}  
i.e. {X<sub>3</sub>, X<sub>4</sub>, X<sub>5</sub>, X<sub>6</sub>, X<sub>7</sub>}

The total cost= 1.118 +3.606 + 1.118 +1.803+0.707=  
8.352

Cost comparison

From step 4 and step 5 we get the total cost are 6.236 and 8.352 respectively. So cost of swapping medoid from Q' to R' is

S= Current total cost – Past total cost

$$= 8.352 - 6.236$$

$$= 2.116 > 0$$

So moving would be a bad idea and the previous choice was a good idea.

Step 6

Select another nonmedoid S'. Let us assume S' = (1.5, 2.0). So now the medoids are C<sub>1</sub> (1, 1) and S' (1.5, 2.0).

Again, calculate distance so as to associate each data object to its nearest medoid. Costs to the nearest medoid are shown bold in the table.

i	C <sub>1</sub>		Data objects (X <sub>i</sub> )		Cost (distance)
3	1.0	1.0	3.0	4.0	$\sqrt{ 1.0 - 3.0 ^2 +  1.0 - 4.0 ^2}$ <b>= 3.606</b>
4	1.0	1.0	5.0	7.0	$\sqrt{ 1.0 - 5.0 ^2 +  1.0 - 7.0 ^2}$ <b>= 7.211</b>
5	1.0	1.0	3.5	5.0	$\sqrt{ 1.0 - 3.5 ^2 +  1.0 - 5.0 ^2}$ <b>= 4.717</b>
6	1.0	1.0	4.5	5.0	$\sqrt{ 1.0 - 4.5 ^2 +  1.0 - 5.0 ^2}$ <b>= 5.315</b>
7	1.0	1.0	3.5	4.5	$\sqrt{ 1.0 - 3.5 ^2 +  1.0 - 4.5 ^2}$ <b>= 4.301</b>

i	S'		Data objects (X <sub>i</sub> )		Cost (distance)
3	1.5	2.0	3.0	4.0	$\sqrt{ 1.5 - 3.0 ^2 +  2.0 - 4.0 ^2}$ <b>= 2.50</b>



4	1. 5	2. 0	5. 0	7.0	$\sqrt{ 1.5 - 5.0 ^2 +  2.0 - 7.0 ^2}$ = <b>6.103</b>
5	1. 5	2. 0	3. 5	5.0	$\sqrt{ 1.5 - 3.5 ^2 +  2.0 - 5.0 ^2}$ = <b>3.606</b>
6	1. 5	2. 0	4. 5	5.0	$\sqrt{ 1.5 - 4.5 ^2 +  2.0 - 5.0 ^2}$ = <b>4.243</b>
7	1. 5	2. 0	3. 5	4.5	$\sqrt{ 1.5 - 3.5 ^2 +  2.0 - 4.5 ^2}$ = <b>3.202</b>

#### Cost comparison

From step 4 and step 6 we get the total cost are 6.236 and 19.654 respectively. So cost of swapping medoid from Q' to S' is

$$\begin{aligned}
 S &= \text{Current total cost} - \text{Past total cost} \\
 &= 19.654 - 6.234 \\
 &= 13.42 > 0
 \end{aligned}$$

So moving would be a bad idea and the choice in step 4 was a good idea. So the configuration does not change after step 4 and algorithm terminates here (i.e. there is no change in the medoids- the medoids are  $X_1$  and  $X_5$ ).

From the step 6, we get the following clusters:

Cluster-1=  $\{(1, 1)\}$  i.e.  $\{X_1\}$

Cluster-2 =  $\{(1.5, 2), (3, 4), (5, 7), (3.5, 5), (4.5, 5), (3.5, 4.5)\}$  i.e.  $\{X_2, X_3, X_4, X_5, X_6, X_7\}$

The total cost=  $2.5 + 6.103 + 3.606 + 4.243 + 3.202 = 19.654$

## VII. COMPARISON RESULTS OF MANHATTAN AND EUCLIDEAN DISTANCE FUNCTION

From the both methods we have seen that the set of clusters are the same and the centroids are  $X_1$  and  $X_5$ .

The following figure is the final graphical diagram for our example that is shown in both steps 4.

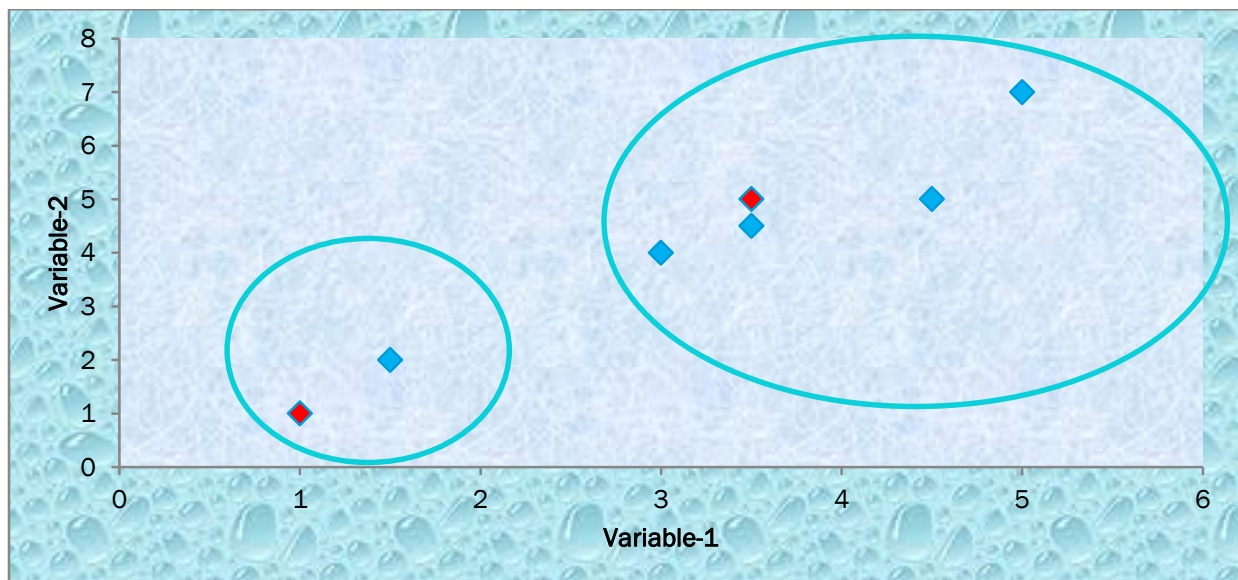


Figure 2: Graphical Diagram of the resultant clustering.

The following figure is the cost function bar-chart diagram for both Manhattan and Euclidean distance.

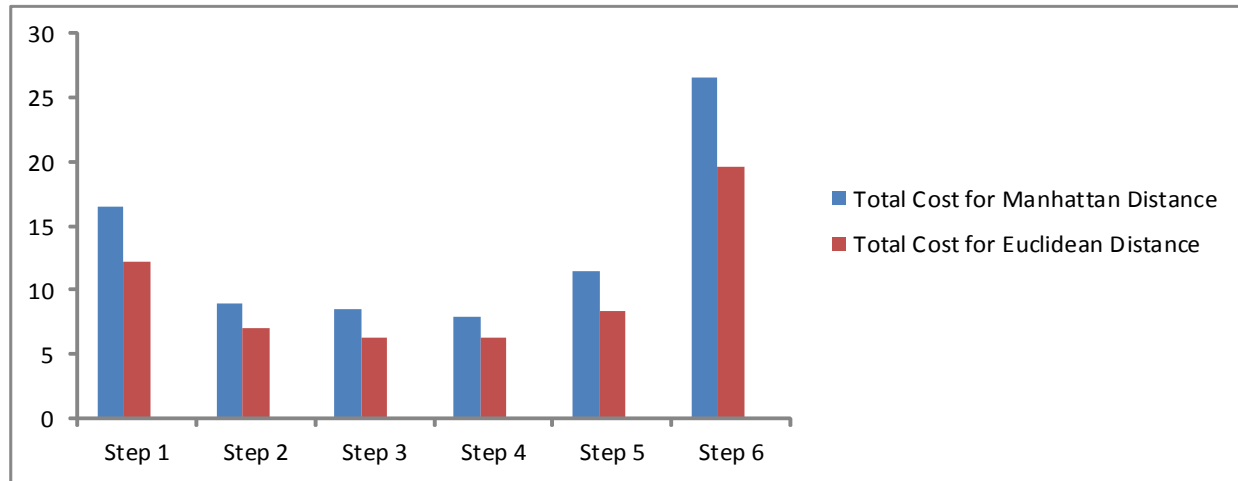


Figure 3: Cost function diagram

We have seen that from the above cost function diagram, the total cost in each step for Euclidean distance is less than the total cost for Manhattan distance. For instance, the total cost in step 1 for Euclidean distance is 12.201 whereas the total cost in the same step for Manhattan distance is 16.5.

#### CONCLUSION AND FUTURE WORKS

Both Manhattan distance function and Euclidean distance function can be used to cluster of data set for the k-medoid. The Manhattan distance is based on absolute value distance, as opposed to squared error (Euclidean) distance. In practice, you should get similar results most of the time. Although absolute value distance should give more robust results, Euclidean distance function is very effective for small amounts of quality data, and thus favor squared error methods with their greater efficiency.

From the comparison result we can deduce that, Euclidean distance function is really effective for a small set of data. In this paper, we have also seen that, the cost function of each step our given example for the k-medoid method using Euclidean distance function is relatively less than the cost function of corresponding step using Manhattan distance function.

In future, I will work on big data set to cluster effectively.

#### REFERENCES

[1] Salissou Moutari, Unsupervised learning: Clustering, Centre for Statistical Science and

Operational Research (CenSSOR), Queen's University, 17<sup>th</sup> September 2013.

[2] Nizar Grira, Michel Crucianu, Nozha Boujemaa. Unsupervised and Semi-supervised Clustering: a Brief Survey in A Review of Machine Learning Techniques for Processing Multimedia Content, Report of the MUSCLE European Network of Excellence (6th Framework Programme), August 15, 2005.

[3] <http://users.ics.aalto.fi/sami/thesis/node9.html>

[4] Shalini S Singh, N C Chauhan. K-means v/s K-medoids: A Comparative Study. National Conference on Recent Trends in Engineering & Technology, 13-14 May 2011.

[5] Amit Singla, Mr. Karambir. Comparative Analysis & Evaluation of Euclidean Distance Function and Manhattan Distance Function Using K-means Algorithm. National Conference on Recent Trends in Engineering & Technology, 13-14 May 2011. IJARCSSE, Volume 2, Issue 7, July 2012.

[6] <http://en.wikipedia.org/wiki/K-medoids>

[7] Deepak Sinwar, Rahul Kaushik. Study of Euclidean and Manhattan Distance Metrics using Simple K-Means Clustering. International Journal for

Research in Applied Science and Engineering Technology (IJRASET) ISSN: 2321-9653, Vol. 2 Issue V, May 2014.

[8] J. Han, M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann Publisher, San Francisco, USA, 2001.

[9] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. From Data Mining to Knowledge Discovery in Databases. American Association for Artificial Intelligence, 1996.

[10] Rahmat Widia Sembiring, Jasni Mohamad Zain, Abdullah Embong. Clustering High Dimensional Data Using Subspace and Projected Clustering Algorithms. International journal of computer science & information Technology (IJCSIT) Vol.2, No.4, August 2010.

[11] Isabelle Guyon, André Elisseeff. An Introduction to Variable and Feature Selection. Journal of Machine Learning Research 3 (2003) 1157-1182.

[12] Charu C. Aggarwal, Jiawei Han, Jianyong Wang, Philip S. Yu. A Framework for Projected Clustering of High Dimensional Data Streams. Proceedings of the 30th VLDB Conference, Toronto, Canada, 2004.

[13] A.K. Jain, M.N. Murty, P.J. Flynn. Data Clustering: A Review. ACM Computing Surveys, Vol. 31, No. 3, September 1999.

[14] Man Lung Yiu and Nikos Mamoulis. Iterative Projected Clustering by Subspace Mining. IEEE Transactions on Knowledge and Data Engineering, Vol. 17, No. 2, February 2005.

[15] Jiawei Han and Micheline Kamber. Data Mining: Concepts and Techniques, Second Edition.

#### AUTHORS PROFILE

1. **Md. Mohibullah** who obtained B.Sc. (Engg.) in Computer Science and Engineering Department at Comilla University, Comilla, Bangladesh. He is now a student of M.Sc. (Thesis) at this university and a member (student) of Bangladesh Computer Society (BCS). His research interest includes data mining, Artificial Intelligent and Robotics.
2. **Md. Zakir Hossain** is now working as Assistant Professor in the Dept. of Computer Science & Engineering at Comilla University, Bangladesh. He was also a former faculty member of Stamford University Bangladesh in the Dept. of Computer Science & Engineering. He obtained MSc and BSc in Computer Science & Engineering from Jahangirnagar University in 2010 & 2008 respectively. His research interest includes Natural Language Processing, Image Processing, Artificial Intelligent and Software Engineering.
3. **Mahmudul Hasan** who obtained an M.Sc. (Thesis) in Computer Science and Engineering from University of Rajshahi, Bangladesh in 2010, is currently employed as Assistant Professor in the Department of Computer Science and Engineering (CSE) at Comilla University, Comilla, Bangladesh. He worked as a Lecturer at Daffodil International University and Dhaka International University in Dhaka, Bangladesh. His teaching experience includes four under graduate courses, as well as five years of research experience at University of Rajshahi, Bangladesh. He is a member of IAENG (International Association of Engineers). His research activities involve Speech Processing, Bio-Informatics, Networking, and Cryptography.

# Proposed GPU Based Architecture for Latent Fingerprint Matching

Yenumula B Reddy

Dept. of Computer Science, GSU, [ybreddy@gram.edu](mailto:ybreddy@gram.edu)

**Abstract**—Most of the fingerprint matching systems use the minutiae-based algorithms with a matching of ridge patterns. These fingerprint matching systems consider ridge activity in the vicinity of minutiae points, which has poorly recorded/captured exemplary prints (information). The MapReduce technique is enough to identify a required fingerprint from the database. In the MapReduce process, minutiae of the latent fingerprint data used as keys to the reference fingerprint database. The latent prints are analyzed using Bezier ridge descriptors to enhance the matching of partial latent against reference fingerprints. We implemented the MapReduce process to select a required document from a stream of documents using MapReduce package. MapReduce model uses parallel processing to generate results. However, it does not have the capability of using Graphics processing units (GPU) to execute faster than CPU-based system. In this research, we proposed a Python based Anaconda Accelerate system that uses GPU architecture to respond faster than MapReduce.

**Key words:** *fingerprint, minutiae points, MapReduce, Bezier ridge, GPU-based architecture;*

## I. INTRODUCTION

Fingerprints help to identify individuals from the unique pattern of whorls and lines. The research concludes that no two individuals (even the twins) have the same fingerprint. The fingerprints do not change one year after the birth of a person. Therefore, the fingerprints are part of the biological data acquisition to identify an individual. The issue raises a new challenge in storing, handling and analysis at the rate it is generated today. Common usage like analysis, comparing, transferring files, and processing is slow and computationally expensive due to the size of the database at any individual workstation. The analysis of fingerprints requires comparison of several features of the fingerprint patterns that include characteristics of ridges and minutia points. Since the amount of fingerprint data is unstructured or semi-structured and increasing exponentially, we need to look for a different solution. The new solution is Hadoop distributed file systems.

Big data is a popular term used for structured, unstructured, and semi-structured large volume of data. Analysis of such data is helpful in business, government and semi-government in operational efficiencies, decision making, reduced risk, and cost reductions. Big data is measured in volume, velocity, and variety. The volume includes the unstructured streaming of social media data,

internet data, e-commerce data, and Government data. The unprecedented speed (velocity) of this volume data must be dealt promptly is a challenging job. These data sets have a variety of formats includes text documents, emails, video, audio, stock data, and financial transactions.

Working with big data does not mean the acquiring of a large amount of data. It is the work you plan to design the unstructured or semi-structured data. The plan is to analyze the data to minimize the cost, real-time response or speed of returning the results, quick decision making, and optimization techniques. The examples of improving the performances are:

- use the new technology to return the real-time response and save dollar amount
- optimize the routes to deliver the goods, analyze the stocks and maximize the profit
- increase the sales based on customer's past purchases
- calculate the risks and protect the business, identify the important customers in related business and
- use the artificial intelligence techniques or use the data mining techniques to improve the business.

The big data analytics considers various types of data to uncover the hidden patterns, unknown correlations, customer preferences, market trends, revenue opportunities and advantageous to respective organizations. Big data strategy can help to pull all related data into a single system to identify the patterns among the customers and identify which type of customers buy a specific product. Big data analysis can be technical, Government, or business related. The organizations use for data management for a quick response. They identified the quick response is through high-performance computing. New algorithms and program techniques are on the way to produce real-time response using high-performance (parallel processing) techniques. The NVIDIA GPU (graphics processing unit) processing helps to achieve the proposed real-time response.

The rate of fingerprint data generated today is very difficult to store and analyze using traditional methods. Therefore, the fingerprint database is one of the largest databases, considered as big data. Further, fingerprint analysis has been used to identify suspects and solve crimes for more than 100 years. The fingerprint identification process to solve the crimes is an extremely valuable tool for law enforcement. Crime scene data (latent fingerprints) is

unstructured data needed to be analyzed and processed before use. Every fingerprint has a unique pattern that appears on the pads of the fingers and thumbs. These unique patterns are made by friction ridges and furrows.

When a crime occurs, law enforcement enables to obtain the fingerprints (latent) from crime area and analyze the patterns and match the potential matches. Currently, the system scans the prints and analyzes all ridges, swirls, loops, and other related patterns to uncover the potential matches. The current matching algorithms do not allow their use in large fingerprint databases due to their computational limitations. The GPU (graphics processing unit) based fingerprint matching methods can overcome these limitations [1 - 2]. Their study shows that GPU-based computation speed up the process and minimizes the cost. Further, the authors suggested that GPU-based processing opens the new field of possibilities to identify and return the possible matching of fingerprints in real-time in large databases.

## II. REVIEW OF WORK

Personal identification is used in many workplaces, passports, cellular phones, credit cards, automatic teller machines, and driver licenses are using the personal identification in an encrypted form. Cryptography plays a significant role to encode and decode the identification for storing and retrieving. In spite of encrypted form, fraud is in credit cards alone reach billions each year in worldwide. Therefore, biometric identification helps to identify an individual in a unique way. Biometric system information verifies and identifies an individual. The biometric information may include voice, face, fingerprints, eyes, retina, signature, keystroke dynamics and much similar identity information. Fingerprints are becoming more common to identify a person and identification machines are becoming cheaper.

Fingerprint matching, authentication, comparative study, the performance of fingerprint quality measures, and the statistical study were done in [3-9]. The fingerprint matching process helps to determine two sets of ridge details come from the same finger. The algorithms use matching minutiae points or similarities between two finger images. Ackerman [10] discussed a method of for fingerprint matching based on minutiae matching. The method uses the preprocessed region analysis to eliminate the processing delay. Bhuyan et al. [11] discussed the fingerprint classification using data mining approach. The proposed approach uses linear k-means robust *apriori* algorithm for the seed selection. The proposed *apriori* algorithm designed on cluster-based model that uses the selected seeds. The authors claim the proposed approach has higher accuracy and eliminates the misclassification errors. Hong and Jain presented fingerprint classification algorithm and tested on NIST-4 fingerprint database [12]. The algorithm classifies the input fingerprints into five categories depending on the number of singular points, their

relative position, and the presence of recurring ridges (type-1 and type-2). The proposed algorithms achieve better performance compared to previous algorithms.

Processing fingerprinting for big data and fingerprint matching problem using NVIDIA GPU are current research in Hadoop distributed file systems using GPU-based implementation. Tretyakov et al. [13] discussed the probabilistic fingerprinting to reduce the use of computational resources and increase in proceeding speed. A Gabor filter bank based algorithm was discussed for fingerprint images using GPU [14]. The authors claimed that GPU-based implementation was 11 times faster than CPU-based implementation. Awan [15] discussed the local invariant feature extraction using graphics processing unit (GPU) and central processing unit (CPU). Their results show that GPU-based implementation return the results faster than CPU-based implementation. The authors used the feature extractors called scale invariant feature transform (SIFT) and speeded-up robust feature (SURF). They concluded that SURF consumes longer matching time compared to SIFT on GPU-based implementation.

In [16], minutia cylinder-code (MCC) based algorithm was used in GPU fingerprint-based system to enhance the performance. The tables show that the GPU-based MCC implementation is 35 times faster than the CPU (single thread) on a single GPU and approximately 100x faster on multiple GPUs. The experiments were conducted on 55k size database. All the above experiments were carried out with known fingerprint search. The papers do not show that the search was carried out with latent prints.

The automated matching of partial latent prints is difficult to current systems. The conventional methods require a sufficient number of ridge bifurcation and termination (minutiae) to support the search. To develop an automated system is a significant challenge to detect a matched fingerprint from the available latent print. This method should not rely on traditional minutiae matching methods. Walch and Reddy [2] used the GPU technology to solve the latent fingerprints matching problem. The authors proposed Bezier curves as ridge descriptors to produce accurate overlays of the latent onto a reference print. The GPU-based method used by the authors performs near real-time results. The comparisons vary from 20 million to 971 billion depending upon the reference Beziers. They claimed that the processing of 8 days on CPU reduced to one hour on GPU cluster.

### Contribution

The research presents the current state of fingerprint algorithms using various techniques that include traditional, pattern recognition, and hybrid methodologies. The Hadoop technology was implemented to the required document identification and then proposed to identify and retrieve the required fingerprint with the latent print. Using this MapReduce technique, we suggested that minutiae data of latent fingerprints can be used as keys to search the

reference fingerprint database. Currently, MapReduce does not work for GPU-based technology for parallel processing, an alternative model was suggested. The new model includes NubmaPro with Anaconda Accelerate allows developers to write parallel code that used NVIDIA GPUs.

### III. FINGERPRINT IDENTIFICATION

Fingerprints are the impressions or mark made on a surface of a person's fingertip. These are unique patterns of recognition of an individual using the pattern of whorls and lines on the finger impression. The combination of ridge bifurcation, trifurcation, bridge, ridge crossing, hook, ridge ending, island (short ridge), dot and similar characteristics determines the uniqueness of the fingerprint. The identification points consist of bifurcations, ending ridges, dots, ridges, and islands. Most of the quality fingerprints contain 60 to 80 minutiae. A single rolled fingerprint may have as many as 100 or more identification points. These points can be used for identification purposes. There is no exact size requirement as the number of points found on a fingerprint impression. The size and latent fingerprint depends on the crime location.

There are many algorithms to match the fingerprints of an individual stored in the database. Some of the algorithms for fingerprint detection include the nearest neighbor, fixed radius, phase-based image matching, feature-based matching and combination of phased-based and feature-based. In the nearest neighbor algorithm, the K adjoining minutiae is considered for matching the neighborhood of known minutiae. The fixed radius algorithm uses a circle of radius centered on minutiae and the neighborhood patterns to match with the database samples. The phase-based image matching algorithm uses the phase components in a two-dimensional (2D) discrete Fourier Transform of the given images. The feature-based matching algorithm uses pairwise corresponding image features of an edge, a corner, a line, or a curve. The combination of phase-based image matching and feature-based matching fingerprint recognition algorithm helps the low-quality fingerprint images.

Fingerprint authentication models include the extraction of raw data, matching the extracted features, get match score and authentication decision. The authentication does not provide complete protection of data. The authentication also depends upon the operating conditions and among individuals. The operating conditions include dirt across fingerprint sensor, malfunction of sensors, and static electricity may cause sensor malfunction. Clency et al. [17] discussed the fundamental insecurities that hamper the biometric authentication and cryptosystem capable of using the fingerprint data with cryptography key. Hong et al. [18] presented the automatics authentication system with fingerprints as an individual identity. Thai and Tam [19] discussed the standardized fingerprint model. Their proposed model synthesizes the template of fingerprints before the process. The steps include preprocessing, adjusting parameters, synthesizing fingerprint and post-

processing. Jain et al. [4] proposed the filter-based fingerprint matching algorithm that uses Gabor filters to capture both local and global details of fingerprints. The matching depends on the Euclidean distance between the two corresponding codes.

Conventional hashing algorithm used for fingerprint matching is an expensive process as the data is increasing in terabytes. The conventional data collectors have different filenames with different data or same content with different file names. If we download the files for processing, most of the times we have duplicate data. Therefore, Tretyakov et al. [13] used an algorithm based on probability theory. The algorithm computes the fingerprint file by using a few data samples. The model can be applied to a network model to retrieve the remote samples faster.

The probability model requires the organized data since hashing is the primary tool to compare and extract. The reference fingerprint data particularly latent prints are unorganized data and need to be analyzed and processed before use. In the analysis, we use particular characteristic called minutiae. Minutiae cylinder-code was used for accurate results [1]. The model is slow due to computational requirement with current CPU speed. The algorithm was modified to use the NVIDIA GPU processors. The performance is many times faster than single CPU, and the problem was to move the data into GPU memory, process and return the results back to host memory to display or return the results. Currently, this process cannot be changed, but in future NVIDIA technology GPUs can access the host memory and process the data at GPUs.

A GPU-based model was also discussed by Walch and Reddy [2] to solve the latent fingerprint problem that is a unique attempt to identify the criminals and terrorists. The authors used handwriting character recognition model. The model uses Bezier descriptors and a means of generating ridge-specific markers. It described through four points that include 2 end points and 2 control points. The model involves finding a subset of corresponding ridge sections common to both latent and reference points. Then it uses the step-by-step curve matching process. The technique also used by Rahman et al. [20] to identify an original image from reference shape. They used the technique called parametric curve generation. The authors compared the results that the curve generated by composite Bezier curve.

For reliable processing, the fingerprint algorithms need to eliminate noise, extract minutiae and rotation and translation-tolerant fingerprint matching. We need to create an algorithm to be suitable for current hardware technology. The hardware may be cloud-based or cloud-based technology using GPUs. The algorithm must complete a large number of comparisons in microseconds. Since current technology with available CPUs meets their limitations, we need to use GPU-based processing to get real-time or near real-time response.

The research focus is on big data analysis and process. We use fingerprint data to store and retrieve. The latent fingerprint data is the search key that is to be analyzed before fingerprint match in the database. Currently, the fingerprint data is outweighing for current computing facilities. Performing the computational analysis of such data into usable information and provide the results in close real-time is the goal. Due to limitations of current CPU, we recommend the GPU technology to process such data. Special algorithms are needed to use GPU technology to meet customer satisfaction.

#### IV. BEZIER CURVE AND FINGERPRINT ANALYSIS

Latent fingerprints are difficult to analyze due to their poor image quality. The extracted latent fingerprint from the crime scene has limitations of minutiae that are required for fingerprint match in a reference database. A latent fingerprint image is normally 20 to 30 percent of an original fingerprint image. Identifying the original fingerprint with 20% of its information and the minimum number of minutiae points required to match a fingerprint is an unsolved problem. Further, the search is a difficult part without enough ridge bifurcations and terminations (minutiae). The latent fingerprints may come from any location (part) of the print. The latent fingerprints may contain meaningful information if they come from the core and less information from other regions of the print. The ridge specific markers help potential information and ridge geometry creates new features to supplement the missing bifurcations and ridge endings. If we create proper ridge specific markers, they should be functionally equal to traditional minutiae.

The third order Bazier curve is a smooth mathematical curve that is used to approximate the curved object such as ridges. Bezier curves precisely fit into the curvature of ridges. These curves are used to mark the positions on the ridges that create 'minutiae'. In Figure 1, the curve consists of two endpoints and two control points. The four related points called endpoints ( $P_0, P_3$ ) and control points ( $P_1, P_2$ ). The control points ( $P_1, P_2$ ) define the shape of the curve. The Bezier curve does not pass through the control points. The control points showed the direction of the curve and positioned outside the curve.

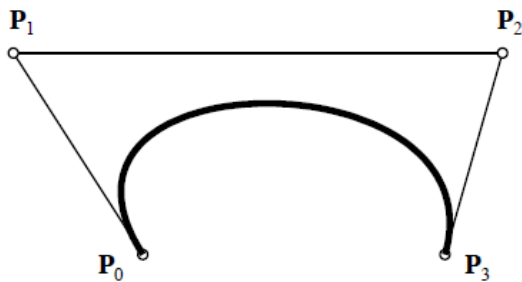


Figure 1: Cubic Bezier Curve

Bezier curves are polynomials of  $t$  that varies 0 to 1. The equation of cubic form of Bezier curve is of the Binomial form with end point  $n$  is given below.

$$Bezier(n, t) = \sum_{i=0}^n \binom{n}{i} (1-t)^{n-i} t^i \quad (1)$$

In the equation (1)

$\binom{n}{i}$	Binomial term
$(1-t)^{n-i} t^i$	Polynomial term
$\Sigma$	series of additions

Bezier curves can be drawn in many ways. Let us consider two ways of Bezier curves. First, use the de Casteljau's algorithm [21]. Consider various points between beginning and end of the curve ( $t=0$  as starting point and  $t=1$  as the end point) with increment  $\Delta t$ . The smooth curve will be obtained with more points between beginning and ending (smaller value of  $\Delta t$ ). The second method is sampling the curve at certain points and joining those points up with straight lines and smoothen the curve along those lines. The disadvantage is that we lose the precision of working with a real curve. We suggest a tool with ridge specific markers (that generates from minutiae) in the lines of handwritten character recognition. The proposed tool becomes a unique in fingerprint recognition using latent prints.

The sampling model can be used for latent Bezier curve creation. Sample each edge in the latent fingerprint skeleton to form the Bezier sample to compare with reference print. The process continues till the minimum acceptance (threshold) point reached between latent prints and reference prints. The matching does not include complete curve matching. The comparison is the similarity between the latent and reference prints. Exact length never happens with a latent print comparison. If the latent fingerprint matches the part of the reference curve, we will continue to match next warp to match and continue to all warps of latent fingerprints. The threshold is set the certain percent of matching with reference print.

#### V. GPU-BASED APPROACH

As the data scales in size, exponential problems of storage, organization, management and retrieval are immediate problems to consider. The new method using GPUs helps to process such data and respond near real-time. Most of the times, GP-GPU cluster is suggested to process petabytes of data in seconds. It is estimated that a system with 16 GPUs performs 160 Trillion calculations per GPU per Hour (2.5 Quadrillion calculations per hour) [2]. In such environment, the data process takes 200 hours can be completed in one hour.

As the data is increasing exponential, a GPU-based process can be used to produce fast results as quickly as MapReduce. The analysis and algorithm for a GPU-based



system are required. In both cases, the following points are important.

- The fingerprint size
- The curvature segment size selected for comparison
- Sample the number of curves using an “Overlapping Stepping” method from ridgelines (Section IV discusses the Casteljaou’s algorithm and sampling latent prints, and reference Bezier curve creation)
- The ratio between latent print to reference fingerprints is 100.

Walch and Reddy [2] provided the calculations for a sample 3 tuple Bezier curve that is set from latent print without indexing. In their calculation, each one of these 3-tuple Bezier curve sets need be compared with reference database to find a “best possible matching” of 3-tuple Bezier curve. The fact is that the curve is fixed in every reference print in the corpus. The calculations provided below are for the comparison of one full print against a database of 100 million reference sets (10 fingerprints). For example, 20k graphs \* 10prints \* 100M sets = 20 trillion comparisons.

The Realistic Metrics utilizing GPU are:

- Combinations for latent print with 500 Beziers  

$${}_{500}C_3 = 500! / 3!(500-3)! = 500! / 3*2*1(497)! = 500*499*498 / 3*2*1 = 124251000/6 = \mathbf{20,708,500}$$
- Combinations for Reference print with 18000 Beziers  

$${}_{18000}C_3 = 18000! / 3!(18000-3)! = 18000! / 3*2*1(17997)! = 18000*17999*17998 / 3*2*1 = 5831028036000/6 = \mathbf{971,838,006,000}$$
- The similarity score (Compare and calculate) for 20,708,500 latent combinations has 971,838,006,000 reference combinations.

## VI. MAPREDUCE METHOD

Big Data processing and analysis uses the MapReduce technology. Big Data is the data that is unstructured, semi-structured, or a combination. The e-commerce data, Facebook data, Twitter data, or any similar data are examples of Big Data. These databases are increasing exponentially and piled up in petabytes. Fingerprint data can be considered as Big Data since it is stored as semi-structured and keywords based upon the latent to search the reference database.

In this paper, the first approach uses the MapReduce techniques. In MapReduce approach, we use latent fingerprint data as keys and retrieve the required fingerprint from the fingerprint database. The approach finds the repetition of each key in each set of fingerprints and retrieves the closest match. The method is similar to many times each keyword repeats in a text file or any stream of data to select particular document is important. The approach is shown in Figure 4.

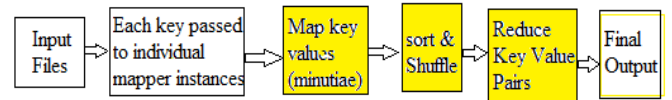


Figure 4: MapReduce Process

To select the required fingerprint file the close matching of laments data is required. Suppose, we set the threshold as 90%, then any fingerprint file accepts this condition will be retrieved.

## VII. MAPREDUCE - TESTING WITH DOCUMENTS

To select a required document Using MapReduce, we provided the keywords and their importance that varies between 0 and 1. We then take the important factor multiplied by the number of times keyword and sum the result of all keyword importance. If the sum is  $\geq$  threshold we conclude that the document is required. The algorithm was coded in two steps. During the first step, the reputation of the words and in the second step the importance factor and selection of the document were coded in Python. We processed six text files to compare the results of the current experiment. The keywords and impact factor provided are: medicine (0.02), profession (0.025), disease (0.02), surgery (0.02), mythology (0.02), and cure (0.05). The size of each file in words, times taken in seconds to process and impact factors respectively are: (128,729; 0.1539; 5.39), (128,805; 0.1496; 0.62), (266,017; 0.13887, 0), (277,478; 0.1692; 6.02), (330,582; 0.1725; 7.93), and (409,113; 0.2032; 18.87). The threshold set was 10.0. Therefore, the file with impact factor 18.87 is selected as required file. If we lower the threshold to 5.0 another two files with impact factors 6.02 and 7.93 would become our required files.

In the proposed fingerprint identification model with MapReduce process, we provide Minutiae data evolved from Bezier curves as keys to search the reference fingerprint database. The present MapReduce package does not have a GPU-based implementation to increase the speed of execution. Therefore, a GPU-based model is required for parallel activity to produce 100X times faster than current existing MapReduce models.

## VIII. ANACONDA FOR BIG DATA ANALYTICS

Analytics comes from Big Data stored in Hadoop and extract values from that information. With Big Data analytics, we can do data mining, text mining, predictive analytics, forecasting, and explore various options to take business decisions. Query, visualize, and perform descriptive statistics

Anaconda Python Accelerate package is created to use CUDA-Python compiler. Math Kernel Library (MKL) is a set of threaded and vectorized math routines that work to accelerate various math functions and applications. Some of the most MKL-powered binary versions of popular numerical and scientific Python libraries are incorporated into MKL for improved performance.

Anaconda Pro is a multi-tool for Big Data. The tool Disco in Anaconda provides Python-based MapReduce Framework. Python is an open source spatial libraries has a facility for Data handling (shapely, GDAL/OGR, pyQGIS, pysnp, pyproj), analysis (numpy, scipy, shapely, pandas, GeoPandas, PySAL, Rasterio, scikit-learn, scikit-image), and plotting (matplotlib, prettyplotlib, Descartes, cartopy).

Anaconda package includes Python (3.4.3, 3.3.5, 2.7.1 and/or 2.6.9) easy to installation and updates of 150 prebuilt scientific and analytic Python packages including NumPy, Pandas, Matplotlib, and IPython with another 340 packages available with a simple “Conda Installation Package Name.”

Accelerate is an Anaconda add-on that allows Python developers to enable fast Python processing on GPU or multi-core CPUs. Anaconda Accelerate designed for large-scale data processing, predictive analytics, and scientific computing. It makes GPU processing easy and an advanced Python for data parallelism. The Accelerate package helps to speed up 20x – 2000x when moving pure Python application to accelerate the critical functions on the GPUs. In many cases, little changes required in the code. The alternate solution to accelerate Python code is PyCUDA that has a capability of calling the CUDA Runtime API. The benefit of PyCUDA is that it uses Python as a wrapper to the CUDA C kernels that develops the programs rapidly.

The NumbaPro comes with Anaconda Accelerate product has CUDA parallel processing capability for data analytics. NumbaPro can compile Python code for execution on CUDA-capable GPUs or multicore CPUs. Currently, it is possible to write standard Python functions and run them on a CUDA-capable GPU using NumbaPro. The NumbaPro package was designed for array-oriented computing tasks. The usage is similar to Python NumPy library. The data parallelism available in NumbaPro has array-oriented computing tasks that are a natural fit for accelerators like GPUs. The benefit of NumbaPro is that it understands NumPy array types to generate efficient compiled code for execution on GPUs. The programming effort minimum and it is as simple as adding a function decorator to instruct NumbaPro to compile for the GPU. Figure 5 shows the GPU-based approach for Big Data analysis using Anaconda Accelerate with NumbaPro (AANP).

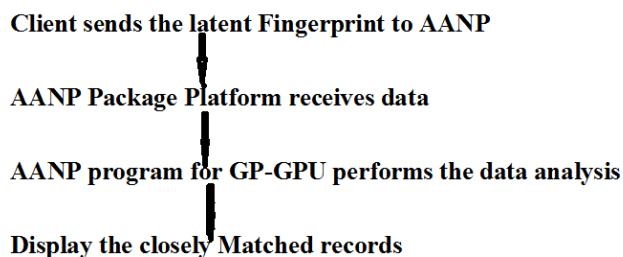


Figure 5: Anaconda Accelerate - GPU Based Approach

We are building the AANP process package to utilize the GPU power and respond quickly. The proposed GPU-

based approach for Big Data analysis includes the GPU/CPU capabilities. The actions include following.

- Preprocess the reference data
- Proposed AANP acts on reference data
- The reference data will then be distributed various nodes in the cluster for processing. In the case of a single system with GPUs, the processing chooses the GPU cluster.
- The processed results will be send back to CPU to display

The proposed AANP is in the process of initial design and uses Python with NVIDIA CUDA capabilities.

## IX. CONCLUSIONS AND FUTURE WORK

The paper discusses the currently available fingerprint identification algorithms, models, the time required to match the key fingerprint with reference print and problems in latent print match matching with reference prints. We need to work in the line of new programming techniques and algorithms for latent print matching using high-performance computing. We tested document identification using Apache MapReduce package. Apache MapReduce does not have capabilities of GPU for analysis of data, so we need alternate approach called GPU-based implementation for fast processing. The GPU-based hardware was suggested to produce near real-time response. We identified that the Python-based NumbaPro with Anaconda Accelerate was more suitable to implement the NVIDIA GPU to analyze the data and expected faster than Apache-based approach.

## REFERENCES

1. P. D. Gutierrez, M. Lastram F. Herrera, and J.M. Benitez., “A high Performance Fingerprint Matching System for Large Databases Based on GPU”, Information Forensics and security, IEEE Transaction on Biometrics Compendium, Vol. 9, Issue 1, 2014, pp. 62-71.
2. “M. A. Walch and Y. S. Reddy., “Using GPU Technology to Solve the Latent Fingerprint Matching Problem,” GTC Express Webinar, July 11, 2012.
3. A. Jain, S. Prabhaka, and A. Ross., “Fingerprint Matching Using Minutiae and Texture Features”, Proceedings of the International Conference on Image Processing (ICIP), Thessaloniki, Greece, 2001, pp. 282–285.
4. A. Jain, S. Prabhakar, L. Hong, and S. Pankanti, “Filterbank-based fingerprint matching,” IEEE Transactions on Image Processing, vol. 9, no. 5, pp. 846–859, May 2000.
5. A. Jain, L. Hong, S. Pankanti and R. Bolle., “An identity authentication system using fingerprints”, Proc. IEEE 85(9), 1365–1388 (1997)
6. F. Alonso-Fernandez, etc., “A comparative study of fingerprint image quality estimation methods”, IEEE Trans. on Information Forensics and Security 2(4), 734–743 (2007)
7. F. Alonso-Fernandez, etc., “Performance of fingerprint quality measures depending on sensor technology”, Journal of Electronic Imaging, Special Section on Biometrics: Advances in Security, Usability and Interoperability (to appear) (2008)
8. A. Bazen and S. Gerez., “Systematic methods for the computation of the directional fields and singular points of

- fingerprints”, *IEEE Trans. on Pattern Analysis and Machine Intelligence* 24, 905–919 (2002)
9. E. Bigun, J. Bigun, B. Duc, and S. Fischer., “Expert conciliation for multi modal person authentication systems by Bayesian statistics”, *Proc. International Conference on Audio- and Video-Based Biometric Person Authentication, AVBPA LNCS-1206*, 291–300 (1997)
  10. A. Ackerman and R. Ostrovsky. "Fingerprint recognition." *UCLA Computer Science Department* (2012).
  11. M. Bhuyan, S. Saharia and D. Bhattacharyya., “An Effective Method for Fingerprint Classification”, *International Arab Journal of e-Technology*, vol.1, no.3, Jan 2010.
  12. L. Hong and A. Jain., “Classification of Fingerprint Images”, 11<sup>th</sup> Scandinavian conf. Image Analysis, Kangerlussuag, Greenland, June 7-11, 1999.
  13. K. Tretyakov, etc., “Fast Probabilistic file fingerprinting for big data”, *BMC Genomics*. 2013, 14 (Suppl 2):S8; ISSN 1471-2164 - p. 8.
  14. R. Lehtihet, W. Oraiby, and M. Benmohammed., “Fingerprint grid enhancement on GPU”, *International conference on Image Processing Computer Vision, and Pattern Recognition (IPCV 2013)*, 2013, pp 1-4.
  15. A. I. Awad., “Fingerprint Local Invariant Feature Extraction on GPU with CUDA”, *Informatica*, vol. 37, 2013, pp. 279-284.
  16. P. D. Gutierrez, M. Lastra, F. Herrera, and J. M. Benitez., “A high Performance Fingerprint Matching System for Large Databases Based on GPU”, *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 1, Jan 2014, pp: 62-71.
  17. T. Clancy, N. Kiyavash and D. Lin., “Secure Smartcard-based Fingerprint Authentication”, *ACM SIGMM workshop on Biometrics methods and applications*, 2003, pp: 45-52.
  18. L. Hong, A. Jain, S. Pankanti and R. Bolle., “Identity Authentication Using Fingerprints”, *First International Conference on Audio- and Video-Based Biometric Person Authentication, (AVBPA) 1997*, pp. 103-110.
  19. L. Thai and N. Tam., “Fingerprint Recognition using Standardized Fingerprint model”, *IJCSI International Journal of Computer Science Issues (IJCSI)*, Vol. 7, Issue 3, No.7, 2010, pp. 11- 17.
  20. M. Rahman, M. Ali and G. Sorwar., “Finding Significant points for Parametric Curve Generation Technique”, *Journal of Advanced Computations*, 2008, vol. 2, no. 2, pp. 107-116.
  21. Farin, Gerald & Hansford, Dianne (2000). *The Essentials of CAGD*. Natic, MA: A K Peters, Ltd. ISBN 1-56881-123-3.
  22. F. Halper, *Eight Considerations for Utilizing Big Data Analytics with Hadoop*, March 2014, SAS report.

# Accelerated FCM algorithm based on GPUs for landcover classification on Landsat-7 imagery

Dinh-Sinh Mai

Le Quy Don Technical University, Hanoi, Vietnam

**Abstract-** Satellite imagery consists of images of Earth or other planets collected by satellites. Satellite images have many applications in meteorology, agriculture, biodiversity conservation, forestry, geology, cartography, regional planning, education, intelligence and warfare. However, satellite image data is of large size, so satellite image processing methods are often used with other methods to improve computing performance on the satellite image. This paper proposes the use of GPUs to improve calculation speed on the satellite image. Test results on the Landsat-7 image shows the method that authors proposed could improve computing speed faster than the case of using only CPUs. This method can be applied to many different types of satellite images, such as Ikonos image, Spot image, Envisat Asar image, etc.

**Index Terms-** Graphics processing units, fuzzy c-mean, land cover classification, satellite image.

## I. INTRODUCTION

Many clustering methods have been proposed by different researchers, especially fuzzy clustering techniques. In recent times, fuzzy clustering methods have been studied and widely used in many applications on the basis of fuzzy theory and the building of the membership function in the range [0..1].

One of the most widely used fuzzy clustering method is the fuzzy c-means (FCM) algorithm [1]. This algorithm was first introduced by Dunn [2] and was later improved by Bezdek [3]. In the FCM algorithm, a data object may belong to more than one cluster with different degrees of membership function. Although the FCM clustering algorithm is popular, its performance is processed slowly on large data sets, many dimensions.

Real-time processing of multispectral images has led to algorithm implementations based on direct projections over clusters and networks of workstations. Both systems are generally expensive. Beside, GPUs are cheap, high-performance, many-core processors that can be used to accelerate a wide range of applications, not only the graphics processing, so we choose the GPUs to solve landcover classification problems on the satellite image [18].

Acceleration problems with satellite images in recent year has achieved quite good results [7], many results have shown that the use of GPUs has significantly reduced processing time.

Anderson et al [4] presented a solution on GPUs for the FCM. This solution used OpenGL and Cg to achieve approximately two orders of magnitude computational speedup for some clustering profiles using an NVIDIA 8800 GPUs. They later generalized the system for the use of non-Euclidean metrics, see Anderson et al [5]. Rumanek et al [16] presents preliminary results of studies concerning possibilities of high performance processing of satellite images using GPUs. At the present state of the study, using distributed GPUs-based computing infrastructure allows to reduce the time of typical computation 5 to 6 times. R.H.Luke et al [17] introduces a parallelization of fuzzy logic - based image processing using GPUs. With results speed improvement to 126 times can be made of the fuzzy edge extraction making its processing realtime at 640x480 image resolution.

A computational speed improvement of over two orders of magnitude, more time can be allocated to higher level computer vision algorithms. Iurie et al [14] presents a framework for mesh clustering solely implemented on the GPUs with a new generic multilevel clustering technique. Chia-F et al [15] proposes the implementation of a zero-order TSK fuzzy neural network (FNN) on GPUs to reduce training time. Harvey et al [6] presents a GPUs solution for fuzzy inference. Moreover, Sergio Sanchez et al [7] used the GPUs to speed up the hyperspectral image processing.

In fact, there are many methods of classifying data, the paper does not mention much about this issue, that only focus research and propose solutions to improve the efficiency of computational for classifying data on a large data, image based GPUs (Graphics Processing Units) but, GPUs architecture not designed for any specific algorithms, with each the algorithm, each data format is designed and installed by programmers. So authors only selected an algorithm to test the problem the research, as the basis for the installation of other classification algorithms on satellite images.

In this paper, we take advantage of the processing power of the GPUs to apply solve the partitioning problem for massive data satellite images based on FCM algorithm. The algorithm must be altered in order to be computed fast on a GPUs.

The paper is organized as follows: Section II shows background; Section III Proposed method, Section IV land cover classification with some experiments; Section V is a conclusion and future works.

## II. BACKGROUND

### A. Graphics processing units

The graphics processing units (GPUs) have become an integral part of today's mainstream computing systems. The modern GPU is not only a powerful graphics engine, but also a highly parallel programmable processor featuring peak arithmetic and memory bandwidth that substantially outpaces its CPU counterpart. Over the past few years, the GPU has evolved from a fixed-function special-purpose processor into a parallel programmable processor by adding an easy-to-use programming interface, which it dubbed CUDA, or Compute Unified Device Architecture [19]. This opened up the possibility to program GPUs without having to learn complex shader languages, or to think only in terms of graphics primitives. CUDA is an extension to the C language that allows GPU code to be written in regular C. The code is either targeted for the host processor (the CPUs) or targeted at the device processor (the GPUs).

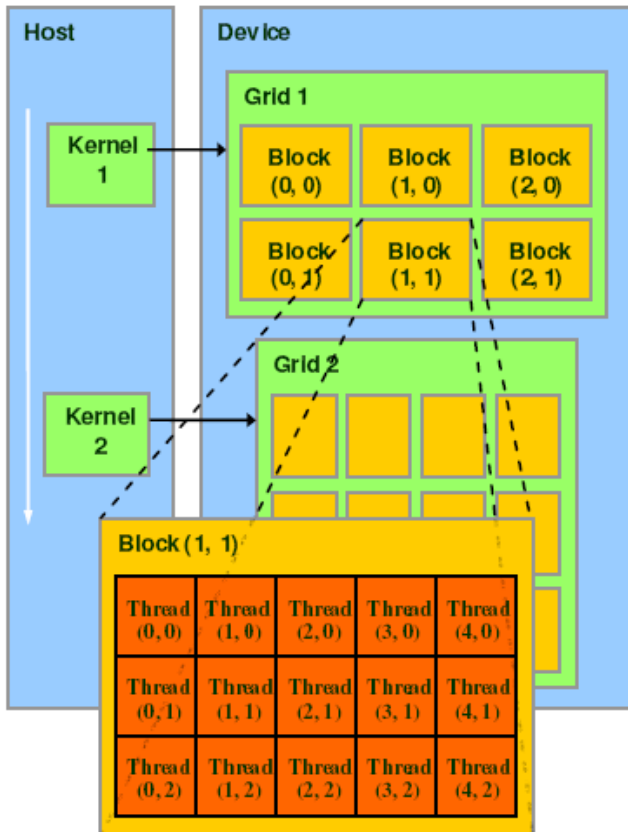


Fig. 1. CUDA GPU's memory model design [19]

CUDA allows multiple kernels to run simultaneously on a single GPU, in which each kernel is a grid. A grid is a collection of blocks. Each block runs on the same kernel but is independent of each other. A block contains threads, which are the smallest divisible units. A thread block is a number of SIMD (Single Instruction, Multiple Data) threads that work on a Streaming Multiprocessor at a given time, can exchange information through the shared memory, and can be synchronized. The operations are systematized as a grid of thread blocks (see Figure 1).

For operation parallelism, the programming model allows a developer to partition a program into several sub-problems, each of which is executed independently on a block. For dataset parallelism, datasets can be divided into smaller chunks that are stored in the shared memory, and each chunk is visible to all threads of the same block. This local data arrangement approach reduces the need to access off-chip global memory, which reduces data access time.

### B. Fuzzy c-means clustering

In general, fuzzy memberships in FCM [1] are achieved by computing the relative distance between the patterns and cluster centroids. Hence, to define the primary membership for a pattern, we define the membership using the value of  $m$ . The use of a fuzzifier gives different objective functions as follows:

$$J_m(U, v) = \sum_{k=1}^N \sum_{i=1}^C (u_{ik})^m d_{ik}^2 \quad (1)$$

In which **Error! Bookmark not defined.** is the Euclidean distance between pattern  $x_k$  and the centroid  $v_i$ ,  $C$  is the number of clusters and  $N$  is the number of patterns. Degree of membership  $u_{ik}$  is determined as follows:

$$u_{ik} = \frac{1}{\sum_{j=1}^C \left( \frac{d_{ik}}{d_{jk}} \right)^{2/(m-1)}} \quad (2)$$

In which,  $i = 1, \dots, C$ ;  $k = 1, \dots, N$ . Cluster centroids are computed as follows:

$$v_i = \frac{\sum_{k=1}^N (u_{ik})^m x_k}{\sum_{k=1}^N (u_{ik})^m} \quad (3)$$

In which,  $i = 1, \dots, C$ . Next, defuzzification for FCM is made as if  $u_i(x_k) > u_j(x_k)$  for  $j=1, \dots, C$  and  $i \neq j$  then  $x_k$  is assigned to cluster  $i$ .

### C. Landsat-7 satellite images

A satellite image consists of several bands of data. For visual display, each band of the image may be displayed one band at a time as a gray scale image, or in combination of three bands at a time as a color composite image. We tested on the Landsat-7 multispectral satellite images with 7 bands and each pixel is a 7-dimensional vector is used for classification in the C class.

The result after classification to classify on the basis of NDVI index (Normalized Difference Vegetation Index), this is the most common measurement to assess the growth and distribution of the vegetation on the earth's surface. Among the seven bands of multi-bands satellite images, using only two bands are NIR (Near-Infrared) and VR (Visible Red) corresponding to band 3 and band 4 in the order of 7 bands, it



has much information about land cover. The NDVI index is calculated as follows:

$$NDVI_{index} = (NIR - VR) / (NIR + VR) \quad (4)$$

Calculations of NDVI for a given pixel always result in a number that ranges from minus one (-1) to plus one (+1); however, no green leaves gives a value close to zero. A zero means no vegetation and close to +1 (0.8 - 0.9) indicates the highest possible density of green leaves. Very low values of NDVI (0.1 and below) correspond to barren areas of rock, sand, or snow. Moderate values represent shrub and grassland (0.2 to 0.3), while high values indicate temperate and tropical rainforests (0.6 to 0.8). For the convenience in processing NDVI data, it is converted to image pixel values and called NDVI image base on the formula:

$$Pixel_{value} = (NDVI + 1) * 127 \quad (5)$$

### III. THE METHOD PROPOSED

#### A. Implementation on the GPUs

To work with GPUs, we need selection of memory types and sizes appropriate [19]. Memory should be allocated such that sequential access (of read and write operations) is as possible as the algorithm will permit. The architecture of a GPUs can be seen as a set of multiprocessors (MPs), the multiprocessors have access to the global GPUs (device) memory while each processor has access to a local shared memory and also to local cache memories in the multiprocessor. In each clock cycle each processor executes the same instruction, but operating on multiple data streams.

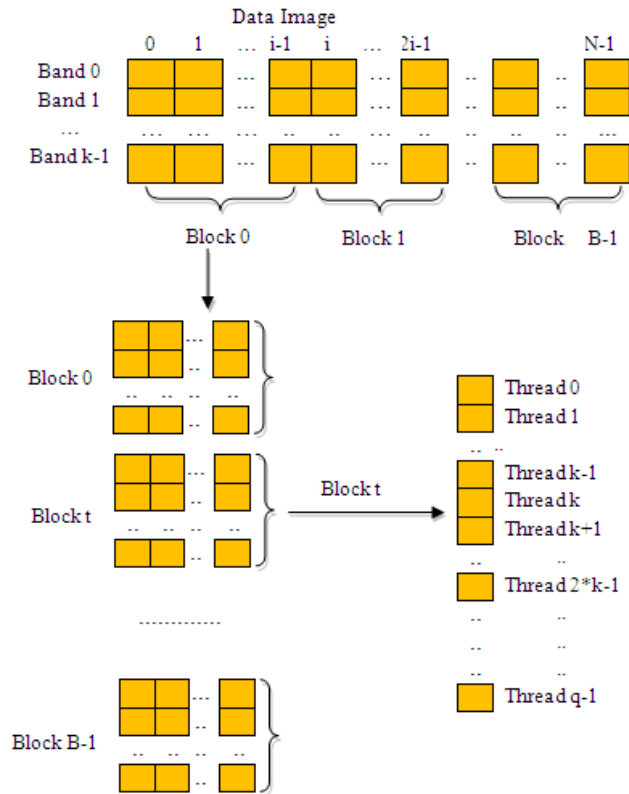


Fig. 2. Data scheme on the GPUs is divided into the blocks and the threads

Algorithms must be able to perform a kind of batch processing arranged in the form of a grid of blocks, where each block is composed by a group of threads that share data efficiently through the shared local memory and synchronize execution for coordinating access to memory. The CPUs should initialize the values for the input array on the GPUs, GPUs calculations give results then returned to the CPUs. CPUs is responsible for showing results.

The inputs from the sample data are of type texture memory because they do not change during the processing. First, we load the Landsat satellite imagery data (X) which has k bands in memory of the CPUs and original initialization C clusters. Satellite imagery is processed with width w, height h, the number of pixels is  $N = w * h$ .

Second, we need copy X and data clusters to the global memory of the GPUs, before to perform the data normalization we use a GPUs kernel, this kernel is configured with as many blocks and maximizing the number of threads per block according to the considered architecture (in our case the number of threads in the block is not greater than 1024 and the number of blocks on the grid is not greater than 65535). Each pixel has k components corresponding to the k bands, so each block is used to calculate  $T = [1024 * k]$  pixels corresponding  $q = T * k$  threads, the number of blocks is used to calculate on the image X is  $B = [N * k / 1024]$ , see Fig.3. Number of clusters is C, so the membership function of the U array of size  $P = N * C$ . The initialization the value of the membership function U corresponding to the fuzzy parameters m. When this processing is completed, we have completed the data normalization to process on the image X.

Third, implement the FCM algorithm. On GPUs, U is calculated simultaneously for the B blocks, B blocks contains N pixels is performed over C clusters and U is calculated by formula 3. With each calculation, check the stop condition, if satisfied copy result to host memory and given the clustering results, otherwise repeat algorithm.

Because there is a limit on the number of threads that can be created for each block (the current maximum of 512 threads per block). So, need to consider the number of threads and registers and local memory requirements by the kernel to avoid memory overflow. This information can be found for each GPUs.

#### B. Land cover classification algorithm

Before the kernel execution, the component means are mapped into the device texture memory (as a k-dimensional CUDA array). These values are cached during the kernel execution. Each thread determines the membership functions with the minimal Euclidian distance between its centroid and the current pixel (each thread operates on one pixel), and stores the index of this component of the membership function matrix. Before executing the kernel, vectors of the C component centroids are copied to the device constant memory. These values are cached once and after wards they are used by each thread from the constant cache, thus optimizing the memory access time. In total  $T = q$  threads are

executed in this task; therefore, the use of the shared memory optimizes the memory access time.

**Algorithm 1: Implementation the FCM on GPUs**

**Step 1:** Initialization data

- 1.1 Reading the Landsat-7 satellite image data into CPUs.
- 1.2 Initialization the parameter of fuzzy m, ( $1 < m$ ), error  $\epsilon$  on CPUs.
- 1.3 Initialization centroid matrix  $V = [v_i]; v_i \in R^d$  by choosing random from dataset on CPUs.
- 1.4 Copy data: satellite image data, m,  $\epsilon$ , V from CPUs memory to GPUs memory.

**Step 2:** Compute the fuzzy partition matrix U and update centroid V on GPUs.

- 2.1.  $j = j + 1$
- 2.2. Fuzzy partition matrix  $U_{ik}$  by the formula 2.
- 2.3. Assign data  $x_j$  to cluster  $c_i$  if data  $(u_{ji} > u_{ki}), k = 1, \dots, c$  and  $j \neq k$ .
- 2.4. Update the centroid cluster  $V_j = [v_{j1}, \dots, v_{jc}]$  by formula 3.

**Step 3:** Check the stop condition on GPUs: Termination criteria is satisfied or maximum iteration is reached, go to step 4, otherwise go to step 2.

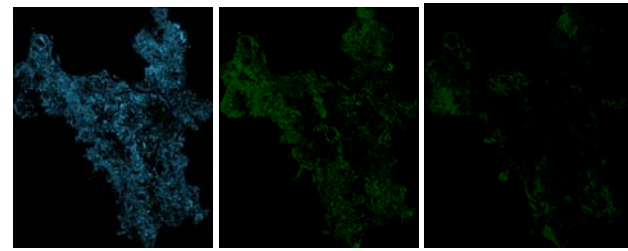
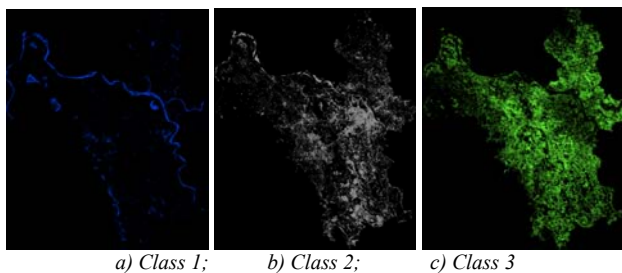
**Step 4:** Compute NDVI index by formula 4 and using NDVI index to assign the corresponding class with 6 types of land cover on GPUs.

**Step 5:** Copy the results and given the clustering results from GPUs memory to CPUs memory.

IV. EXPERIMENTS

The problem was performed on a computer with the operating system was windows 7 64bit and nVIDIA CUDA support with specifications: CPUs was the Core i5-4210U, 2.7 GHz, the system had 6 Gb of system RAM (DDR3). GPUs was an nVIDIA Gerforce GT 750M, graphics card with 384 CUDA Core, 2GB of texture memory.

We test the proposed method with a satellite image LANDSAT-7 taken at Hanoi area, see Figure.5, Hanoi is the capital of Vietnam,  $11^0 24'02.32''N, 107^0 36'26.74''E$  to  $10^0 50'24.61''N, 108^0 09'50.57''E$ , with area is  $3161.304 \text{ km}^2$  and capacity is 187.59Mb.



d) Class 4; e) Class 5; f) Class 6  
Fig. 3: Results classified by 6 classes.

TABLE 1: RESULT OF LAND COVER CLASSIFICATION

Class	N. of pixels	Percentage	Square (hec.)
1	204 892	5.833 %	18439.89
2	690 736	19.665 %	62167.04
3	879 571	25.041 %	79162.21
4	697 474	19.857 %	62774.01
5	621 799	17.702 %	55961.42
6	418 088	11.903 %	37629.71

The results are shown in figure.3 in which (a), (b), (c), (d), (e) and (f) are land cover classification in according 6 class from 1 to 6, respectively. The figure.4 is NDVI image and result image, in which, the NDVI image shows that white areas are rich in vegetation, dark areas are rich in water and result image is synthesized from 6 classes after classification. The table.I shows detailed results according to the number of pixels, the percentage and the area of each class.

Test results Table.I showed that the average processing time on CPUs is 1195.566s, on GPUs is 11.189s and implement time on GPUs faster than on CPUs is 106.852 times. This indicates, GPUs is faster processing and more efficiently than on CPUs.

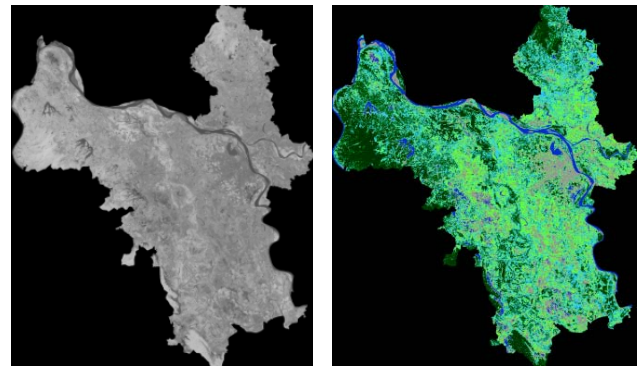


Fig. 4. a) NDVI Image; b) Result Image

To assessing the performance of the algorithms on the test images we compare classification results with statistical data from the Vietnam National Remote Sensing Center (VNRSC) II. The values of these deviations are shown in the Table.II, we noticed the difference between the classes do not exceed 6.08%, in which the deviation at least with 0.428% for class 1 (Rivers, ponds, lakes). Because the area of rivers, ponds and lakes usually have clear boundaries with other classes, therefore resulting classification will give more accurate results than the other classes.

TABLE II: CLASSIFICATION RESULTS OF HANOI AREA

Class	FCM	VNRSC	Deviation
1	5.83%	6.26%	0.43%
2	19.67%	23.65%	3.99%
3	25.04%	31.12%	6.08%
4	19.86%	15.85%	4.01%
5	17.70%	15.16%	2.51%
6	11.90%	7.95%	3.95%

We have tested on several images with different size from small to large. The rate of processing time on CPUs/GPUs shows in Table.III, with image size is 1024x1024, the rate CPUs/GPUs is 41.247 times, when image size is 8192x8192, the rate CPUs/GPUs increases 116.650 times.

This rate is smaller than with smaller size image because the time to read into memory on the GPUs occupies a significant amount. When the size of the image increasing, this rate will be increased. This rate depends also on the characteristics of each area, conditional convergence of the algorithm, the number of loops and so on. Because, it takes time to transfer the data from the CPUs to GPUs and back. When the size of the data increasing, the number of calculations more, the speed of processing on the GPUs faster than on the CPUs. This speed depends on the computer configuration and how to organize the data in the program. With satellite imagery resolution is 30mx30m, the classification results show the highest deviation than 6%. The deviations of classification results can accept in the assessment of land cover on a large area quickly, can allow us to implement large data image processing problems in practice, reduce time and costs compared to other methods. We can improve the accuracy by improving the algorithm or enhance the quality of satellite images before classifying.

TABLE III: CPUs/GPUs PERFORMANCE TIME RATE

Size	1024x1024	2048x2048	4096x4096	8192x8192
GPUs	1.984	3.345	5.986	12.976
CPUs	81.835	216.974	526.462	1513.653
Rate	41.247	64.865	87.948	116.65

## V. CONCLUSION

This paper proposes the method using the GPUs to enhance the computational efficiency for landcover classification problem on the Landsat-7 image based the FCM algorithm. The experimental results show that the implementation is much faster than the traditional implementation on CPUs We have demonstrated that with size and capacity of image large such as satellite images, the processing time on GPUs is much faster than on the CPUs.

The next goal is to implement further research on GPUs for hyper-spectral satellite imagery for environmental classification, assessment of land surface temperature changes. Improved algorithms and data optimized for this purpose.

## REFERENCES

[1] James C.Bezdek, Robert Ehrlich, William Full. "FCM: The Fuzzy C-Means clustering algorithm". Computers and Geosciences Vol.10, No.2-3, 1984, 191-203.

[2] J. C. Dunn, "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters", Journal of Cybernetics 3, 1973, 32-57.

[3] J. C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms", Plenum Press, New York, 1981.

[4] Anderson, D., Luke, R., Keller, J.: *Speedup of Fuzzy Clustering Through Stream Processing on Graphics Processing Units*, IEEE Trans. on Fuzzy Systems, Vol.16:4, pp. 1101- 1106 (2007).

[5] Anderson: *Parallelisation of Fuzzy Inference on a Graphics Processor Unit Using the Compute Unified Device Architecture*, The 2008 UK Workshop on Computational Intelligence, UKCI 2008, pp 1-6 (2008).

[6] Harvey, N., Luke, R., Keller, J., Anderson, D.: *Speedup of Fuzzy Logic through Stream Processing on Graphics Processing Units*, Proc. IEEE Congress on Evolutionary Computation (2008).

[7] Sergio Sanchez, Rui Ramalho, Leonel Sousa, Antonio Plaza.: *Real-time implementation of remotely sensed hyperspectral image unmixing on GPUs*, Proc. J Real-Time Image Proc (2012).

[8] Zhong-dong Wu, Wei-xin Xie and Jian-ping Yu, *Fuzzy C-Means Clustering Algorithm based on Kernel Method*, Proceedings of the 5<sup>th</sup> International Conference on Computational Intelligence and Multimedia Applications, pp.1-6, IEEE 2003.

[9] D.Q. Zhang and S.C. Chen, *Clustering incomplete data using kernel based fuzzy c-means algorithm*, Neural Processing Letter, vol. 18, no. 3, pp. 155-162, 2003.

[10] D. Graves and W. Pedrycz, *Kernel-based fuzzy clustering and fuzzy clustering: A comparative experimental study*, Fuzzy Sets and Systems, vol. 161, no. 4, pp. 522-543, 2010.

[11] R. Hathaway, J. Huband, and J. Bezdek, *A kernelized non-euclidean relational fuzzy c-means algorithm*, 14th IEEE Int. Conference on Fuzzy Systems, pp. 414-419, 2005.

[12] J.-H. Chiang and P.-Y. Hao, *A new kernel-based fuzzy clustering approach: Support vector clustering with cell growing*, IEEE Trans. Fuzzy Systems, vol. 11, no. 4, pp. 518-527, 2003.

[13] C. Yu, Y. Li, A. Liu, J. Liu, *A Novel Modified Kernel Fuzzy C-Means Clustering Algorithm on Image Segmentation*, the 14th International Conference on Computational Science and Engineering (CSE), pp.621-626, 2011.

[14] Iurie, C., Andreas, K.: *GPU-Based Multi level Clustering, Visualization and Computer Graphics*, IEEE Trans on, vol.17,no.2, pp.132-145 (2011).

[15] Chia-F., Teng-C., Wei-Y.: *Speedup of Implementing Fuzzy Neural Networks With High-Dimensional Inputs Through Parallel Processing on Graphic Processing Units*, Fuzzy Systems, IEEE Trans on, vol.19,no.4, pp.717-728 (2011).

[16] Rumanek, Danek and Lesniak.: *High Performance Image Processing of Satellite Images using Graphics Processing Units*, Geoscience and Remote Sensing Symposium (IGARSS), IEEE International, pp 559-561 (2011).

[17] R.H.Luke, D.T.Anderson, J.M.Keller and S.Coupland.: *Fuzzy Logic Based Image Processing Using Graphics Processor Units*, IFAEUSFLAT (2009).

[18] Rauf K. S., Valentin V. G., Leonid P. P.: *Fuzzy clustering methods in Multispectral Satellite Image Segmentation*, International Journal of Computing, Vol. 8, Issue 1, 87-94 (2009).

[19] NVIDIA CUDA, <http://www.nvidia.com>.





**About the Author:**

**Dinh-Sinh Mai** received the B.S. degree in Information Technology and M.S. degree in computer science from Le Quy Don Technical University (LQDTU), Hanoi, Vietnam in 2009 and 2013, respectively.

Now, he is a PhD student and a lecturer at the LQDTU. His research interests involve image processing, fuzzy clustering, pattern recognition and geographic information systems (in computer science).

Phone: (+84)988.906.223

Email: [maidinhsinh@gmail.com](mailto:maidinhsinh@gmail.com)

# Object Oriented Software Metrics for Maintainability

Mr. N.V.Syma Kumar Dasari  
Research Scholar, Dept. of Computer Science,  
Krishna University, Machilipatnam, A.P., India.

Dr. Satya Prasad Raavi  
Assoc. Professor, Dept. of CSE,  
Acharya Nagarjuna University, Guntur. A.P., India.

**Abstract** - Measurement of the maintainability and its factors is a typical task in finding the software quality on development phase of the system. Maintainability factors are understandability, modifiability, and analyzability...etc. The factors Understandability and Modifiability are the two important attributes of the system maintainability. So, metric selections for the both factors give the good results in system of maintainability rather than the existed models. In the existing metrics obtained for Understandability and Modifiability factors based on only generalization (inheritance) of the structural properties of the system design. In this paper we proposed SatyaPrasad-Kumar (SK) metrics for those two factors with help of more structural properties of the system. Our proposed metrics were validated against the Weyker's properties also and got the results in good manner. When compare our proposed metrics are better than the other well-known OO (Object-Oriented) design metrics in getting the Weyker's properties validation.



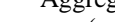

**Keywords** – Understandability; Modifiability; Structural metrics; System Maintainability;; Weyker's properties; SK metrics; OO design;

## I. INTRODUCTION

Maintainability is the most important attribute rather than other attributes like Portability, Usability, and Functionality...etc., for a good quality product as per the ISO-9126 standard in the development phase of the software. Several studies [22], [25] were inexistence in the improvement of the software quality by different researchers. Software maintainability is a dependent factor on the fields of Understandability, Modifiability, Analyzability, Reusability, and Durability...etc., [9],[14].The ISO/IEC9126-1[17] standard gave considerable definitions for the factors of Maintainability i.e., Understandability and Modifiability.

In this paper we developed the metrics for the understandability and modifiability, which plays

the effective role in the process of finding the maintainability of the OO software system. The observed data [20] states that not only inheritance but also other structural properties were played major role in the selection of the metrics for understandability and modifiability. In the design phase of the any OO design software system structural properties plays the vital role. Hence as per our study consideration of all the structural properties in the process of developing the metrics for maintainability factors would give the most prominent results.

The structural metrics were developed on the basis of four fields. The first one is associations (Number of Associations(NAssoc.)), Second one is aggregations (Number of Aggregations(NAgg), Maximum Hagg (Max Hagg) Number of Aggregation Hierarchies (NaggH)),Third one is dependency(Number of Dependencies(NDep)) and the last one is generalization (Number of generalizations(NGen),Maximum DIT (Max DIT), Number of Generalizations Hierarchies (NGenH)).The figures which are shown in Appendix-A reveals the different relationships namely Association(  ), Dependency (  ), Aggregation(  ) and Generalization (  ). Here Generalization field can behave as inheritance in the OO designs. The Super Classes (Sup-C) and Sub Classes (Sub-C) are taken from the generalization field for this research paper. For the remaining all classes which have dependency, aggregation and association are taken as the Connected Classes (Con-C) in our metrics section in this paper.

Every software metric has to show its mathematical and theoretical background by fulfilling the well-known properties suggested by weyker[24] for developing the good software metrics. Many more inheritance based metrics [3], [4], [7], [8], [15],

[18],[19] were evaluated by different researchers against the weyker's proposed rules. Some of the researchers [7],[10],[23] questioned the validity of the weyker's rules because some of the properties were not suited for all types of programming languages.

This paper is organized in following manner. In Section1 we discussed the basic information regarding the research topic. Section2 deals with related work in this research area. Proposed metrics were discussed in effective manner in Section3. Section4 reveals the validity of the proposed metrics against the weyker's properties. Comparison with other class oriented metrics were represented in Section5. Conclusion and Future work regarding our research work was placed in Section6.

## II. LITERATURE SURVEY

Maintainability and its factors like understandability, modifiability, analyzability, portability, usability ..... etc., are the non-functional requirements for the system. Hence so many authors designed the various models [2],[5],[6],[11],[12], for identifying the dependent factors like understandability, modifiability, maintainability.....etc. The metric selection for these types of non-functional requirements is a herculean task in the software system. In the Object-Oriented design of the system so many metrics were developed by different scientists for improving the understanding capability of the given system design.

Depth-of-Inheritance (DIT) and Number of Childs (NOC) metrics were developed by Chidember-Kerner [1],[15] state that maximum depth from the root node to the present node. Here in the DIT technique ambiguity may raise in several situations. In the NOC metric Chidember-Kerner mainly focused on the inheritance hierarchy instead of depth. NOC states that how many immediate sub classes are existed for the individual class. Here the problem is not focusing on total classes at a time. W.Li [13] presented two more new metrics for solving the problems raised by the metrics founded by Chidember-Kerner. The first one is Number of Ancestor Classes (NAC) states that the number of classes inherited by the individual class in the OO design. The second metric is Number of Descendant Classes (NDC) consider the total number of sub classes into the account.

The Average Depth of Inheritance (AID) metric was developed by Henderson-Sellers [21] for applying the average complexity values in the DIT metric. AID metric states that division of sum of depths for the individual nodes in the system with the total number of nodes in the system design. This

metric gives the good results but this may take more time for identifying the metric value in some OO designs.

In the process of development of the metrics for Understandability and Modifiability Sheldon and Jerath [16] proposed the metrics named as Average Understandability(AU) and Average Modifiability(AM) for finding the system understandability and modifiability with considering only inheritance of the OO class diagrams. Here AU metric focused on the super classes (predecessors) of the given individual class only. AM metric consider the AU and sub classes (Successors) of the given class. In these two metrics the authors only focused on the inheritance (Generalization) factor. In the structural representation of the given OO design of the system was represented with associations, dependencies, aggregations and generalizations of the classes. Hence we take all the four factors into the consideration of finding the metrics for the Understandability and Modifiability. In our proposed metrics also we had given the participation to all the four factors in equal manner to identify the Understandability and Modifiability of the OO system in good manner.

## III. PROPOSED METRICS

In this paper we proposed two metrics on the name of SatyaPrasad- Kumar (SK) metrics for identifying the Understandability of the OO system called System Understandability (SU) and another metric for the purpose of modifiability of the system called System Modifiability(SM).

The first metric from the SK metrics is System Understandability (SU) of the OO design means the metric must consider the all the transactions relating to the individual given class. The transactions of the class may be any of the four factors namely associations, dependencies, aggregations and generalizations. In the process of improvisation of the understanding capability of the individual class must consider all the immediate connections of the associations, dependencies and aggregations. In the generalization (Inheritance) phase the data of the given individual class may be utilized in the sub classes of the prescribed class. Hence both sub and super classes must taken into the consideration to find the understandability. For better results of the Understandability metric the prescribed class also has take into the account.

Individual class Understandability (ICU) is as follows

ICU of a class  $ICU_i = Sup-C_i + Sub-C_i + Con-C_i + 1$

$ICU_i$  is the ith class Understandability.

Sup-C<sub>i</sub> is the Number of Super Classes for the i<sup>th</sup> class.

Sub-C<sub>i</sub> is the Number of Sub classes of the i<sup>th</sup> class.

Con-C<sub>i</sub> is the Number of Immediate Connected classes of the i<sup>th</sup> class

SU of the system =

$$\sum_{i=1}^n (\text{Sup-C}_i + \text{Sub-C}_i + \text{Con-C}_i + 1) / n$$

The second metric from the SK metrics is System Modifiability(SM) of the class oriented system design. The SM metrics states that before knowing the information about any class whether it was modified or not we must understand the class first. Then we can focus on the class diagram for modification how many classes modified with modifying the individual class. In this process of modification we must consider the two fields. One is Generalizations because property of the inheritance is modification of one class means total sub classes of that particular class may be modified. Second one is dependency also causes the modification of the class because one class depend on the another class with dependency field. Hence we must consider the generalization and dependency fields also for modification of the system along with the understandability. Generalization (Inheritance) field modification would be applicable to the sub classes of the given class. By considering average case modification half of the subclasses need to modified when modifying one class.

Individual class Modifiability (ICM) is as follows

$$\text{ICM of a class } \text{ICM}_i = \text{ICU}_i + (\text{Sub-C}_i / 2) + \text{ND}_i$$

ICM<sub>i</sub> is the i<sup>th</sup> class Modifiability.

Sub-C<sub>i</sub> is the Number of Sub classes of the i<sup>th</sup> class.

ND<sub>i</sub> is the Number of the Dependencies of the i<sup>th</sup> class.

SM of the system =

$$\text{SU} + \sum_{i=1}^n ((\text{Sub-C}_i / 2) + \text{ND}_i) / n$$

#### IV. VALIDATION OF PROPOSED METRICS

The statistical evaluation of the software metrics can be done against the satisfaction of the weyker's properties leads to good metrics for future use. Here our proposed metrics based on the OO class design diagrams not the inside data and methods of the class. Hence here also some of the weyker's properties(7,9) were not suited for our

proposed metrics which were already not suited for the well-known metrics like DIT, NOC, NAC, NDC, AID, AU, AM because those metrics also developed based on the classes not the inside information of the classes.

*Property-1: Non-Coarseness:*

The class A and class B must show the different metric values mean  $M(A) \neq M(B)$ . The figure-1 from the Appendix-A shows the different metric values for different classes. Hence weyker's property-1 Non Coarseness is satisfied by our proposed SU and SM metrics.

*Property-2: Granularity:*

This property requires that the same metric value pose by different cases. Finite set of applications deals with the finite set of classes, hence this property satisfied by any metric designed at class level [19]. Here our proposed SU and SM metrics were also developed based at the class level. Hence our proposed metrics also satisfied the Granularity property.

*Property-3: Non-Uniqueness (Notion of Equivalence):*

This property states that equal complexity values shown by the two different classes A and B for the given metric mean  $M(A) = M(B)$ . The two different classes have the equal metric value. The figure-1 from the Appendix-A shows that same metric result given by the different classes. Hence our proposed SU and SM metrics were also satisfies the Non-Uniqueness property successfully.

*Property-4: Design details are Crucial:*

The property-4 specifies that same function performed by two different designs but the metric value not giving equal result. Suppose class A and Class B designs are different but functions of the designs are same  $M(A)$  is not equal to the  $M(B)$ . our metrics SU and SM are design implementation dependent means for the different designs they give the different metric values. Hence our proposed metrics were satisfies the weyker's fourth property.

*Property-5: Monotonicity:*

This property states that combination of two different classes metric value must be greater than or equal to the individual classes. Suppose Class A and Class B are the two different classes the metric value of the combination classes at least  $M(A+B) \geq M(A)$  and  $M(A+B) \geq M(B)$ .

Here three possible cases must be existed.

- (i) When class A and class B are siblings

As per our proposed SK metrics the figure-2 of Appendix-A shows that class 17 is having the ICU is 5 and ICM is 6. Class 18 shows the ICU is 5 and ICM is 6. When combined both the classes (17+18) gives the ICU value 6 which is greater than the individual classes 17 and 18. ICM value of the class (17+18) is 8 which is greater than the individual metric values of the classes 17 and class 18. So  $M(A+B) \geq M(A)$  and  $M(A+B) \geq M(B)$  condition was satisfied in this situation.

Hence first case of the Property-5 was satisfied by our proposed metrics.

- (ii) When one class A is child of another class B.

Consider the figure-3 of Appendix-A says that class 16 is having the ICU value is 6 and ICM value is 8. Class 17 has the ICU value as 5 and ICM value as 6. When combining the both of the classes as class (16+17) gives the ICU value is 8 and ICM value is 10.5. The ICU and ICM metric values of the class (16+17) greater than the metric values of the individual classes. Here also  $M(A+B) \geq M(A)$  and  $M(A+B) \geq M(B)$  condition was satisfied successfully.

Hence second case of the Property-5 was satisfied by our proposed metrics.

- (iii) When class A and B are neither siblings nor children of each other.

As per shown in figure-4 of Appendix-A states that class 5 is having the ICU metric value is 3 and the ICM metric value is also 3. Class 9 shows the ICU value is 6 and ICM value is 7.5. The combination of the both classes class (5+9) gives the ICU value is 8 and the ICM metric value is 9.5. The ICU and ICM metric value of the class (5+9) is greater than the metric values of the individual classes. The  $M(A+B) \geq M(A)$  and  $M(A+B) \geq M(B)$  condition was satisfied.

Hence Third case of the Property-5 was satisfied by our proposed metrics.

The property-5 of weyker's was also satisfied by our system level SK metrics (SU & SM) by consider the two system designs and then combining the two designs.

#### *Property-6: Non-equivalence of Interaction*

This property-6 states that if class A and class B shows the equal metric values but the interaction of the other class C with these both of the classes individually need not be equal.

If  $M(A)=M(B)$  but not satisfy that  $M(A+C)=M(B+C)$

The figure-1 of the Appendix -A shows that class17 and class18 have the equal ICU metric value 5 and also equal ICM metric value 6. The combination of class16 and class17 leads to figure-5 of Appendix-A gives the ICU metric value is 8 and ICM metric value 12. Class-18 and class-16 combination shown in figure-6 of Appendix-A gives the resultant ICU metric value is 6 and ICM metric value 7.5. So the ICU and ICM metric values of the class (16+17) are not equal to the metric value of class (16+18). Hence the ICU metric satisfies the Property-6 of weyker's successfully.

The weyker's property-6 was also satisfied by our proposed SK metrics (SU&SM) by adding the new system design to the two equal metric existing system designs individually leads to the different results.

#### *Property-7: Importance of Permutation*

This property-7 states that in the process of permutation of the program statements the metric values of the programs can be changed. This property applicable in traditional programming which can utilize the inside data of the program like as order of the if – else blocks can show the significant effect change of program logic. This property not applicable to OOD metrics suggested by cherniavsky and smith[7].our proposed SK metrics were also based on the OO design .Hence SU and SM metrics do not satisfy the weyker's property-7.

#### *Property-8: Renaming Property*

This property states that if name of the entities changes the metric values of the entities need not changed. Our proposed metrics are based on OO design means we take the class names as entities. If class names are change the values will not change in our SK metrics because our metrics are not depending on the class names. Hence our proposed SK (SU & SM) metrics satisfies the weyker's property-8 in effective manner.

*Property-9: Increased Complexity with Interaction*

This property-9 states that the addition of the two individual classes class-A and class-B metric values are less than or equal to the metric values of the combination of two classes i.e.

$$M(A) + M(B) \leq M(A+B)$$

As per shown in weyker's property-5 with three cases this property-9 also must satisfy the all cases. The weyker's property-9 is not suitable for structural inheritance metrics[3],[18].This property is not satisfied by our proposed metrics because in the class diagram oriented designs gives the metric values of two combined classes is slightly greater or equal to the individual class metric values. In any case the addition of the individual class metric values not less than or equal to the combined classes metric value. Hence our proposed SK metrics (SU &SM) are not satisfies the weyker's property-9.

V. COMPARISON WITH OTHER METRICS

Here our proposed metrics SU and SM for the Understandability and Modifiability are compared to existing proven and well-known metrics. The comparison of our metrics with the metrics named as DIT, NOC, NAC, NDC, AID, AU and AM. The reason behind the comparison is our metrics were also developed based only on the OO class design not focused on the inside data of the classes. Here the comparison table was given below.

TABLE I: MEASUREMENT OF OO METRICS IN VIEW OF WEYKER'S PROPERTIES.

	D I T	N O C	N A C	N D C	A I D	A U	A M	S U	S M
1	√	√	√	√	√	√	√	√	√
2	√	√	√	√	√	√	√	√	√
3	√	√	√	√	√	√	√	√	√
4	√	√	√	√	√	√	√	√	√
5	×	√	×	×	×	×	×	√	√
6	√	√	√	√	√	√	√	√	√
7	×	×	×	×	×	×	×	×	×
8	√	√	√	√	√	√	√	√	√
9	×	×	×	×	×	×	×	×	×

√ - weyker's property satisfied by the metric.  
× - weyker's property not satisfied by the metric.

From the above table shown OOD metrics we found that some properties of weyker's were not satisfied because those metrics were not suited for class level design metrics in OO paradigm. Our proposed metrics are also not satisfies the weyker's(7,9) properties as per all the well-known metrics remaining 7 properties of weyker's satisfied by our proposed metrics in effective manner.

VI. CONCLUSION & FUTURE WORK

Software maintainability shows the considerable effect on the software quality. Software maintainability depends on the understandability and modifiability factors rather than other factors. In this paper we proposed SK(SU & SM) metrics for calculating the understandability and modifiability of the system. Here we observed that the understandability and modifiability of a system in OOD must depend on the structural properties of the system. These structural properties are not only generalizations (Inheritance) but also associations, dependencies and aggregations of the system design. Here we derived individual metrics for understandability (SU) and modifiability(SM) and validated with well-known weyker's properties. In the validation process with weyker's metrics also we got the good results compared other well-known metrics earlier existed.

We already derived metrics for only two factors of maintainability (i.e., understandability and modifiability) with structural properties of the system. We observed that Analyzability also depend on the structural properties of the system. In future we want to concentrate on the structural properties lead to give the metric for the analyzability which is one of the important factors of the maintainability. We want to focus on the dependency, aggregation and association factors of structural properties of the system those may show the considerable effect on the system understandability. With the help of this understandability, modifiability and analyzability metrics we want get the effective results for system maintainability compared with developed models using regression process.

REFERENCES

[1] Chidamber, S.R.—Kemerer, C.F.: Towards A Metrics Suite for Object Oriented Design,OOPSLA'91, pp. 197-211,1991.  
[2] D.N.V.Syma Kumar, R.Satya Prasad and R.R.L .Kantam "Maintainability of Object-Oriented Software Metrics Using Non-Linear Model "International Journal of Advanced Research in Computer Science Engineering and Information Technology Volume: 5 Issue: 3 20-Mar-2015..  
[3] Sharma, N.—Joshi, P.—Joshi, R.K.: Applicability of Weyker's Property 9 to Object-Oriented Metrics. IEEE Transaction on Software Engineering, Vol. 32, 2006, No. 3, pp. 209-211.  
[4] K. Rajnish and V. Bhattacharjee, "Class Inheritance Metrics-An Analytical and Empirical Approach", *INFOCOMP-Journal of Computer Science*, Federal University of Lavras, Brazil, Vol. 7 No.3, pp. 25-34, 2008.  
[5]R.Satya Prasad and D.N.V. Syma Kumar "Maintainability of Object-Oriented Software Metrics with Analyzability" International Journal of Computer Science issues, Volume12,Issue3,May 2015.  
[6] S. Muthanna, K. Kontogiannis, K. Ponnambalam, and B. Stacey, "A Maintainability Model for Industrial Software Systems Using Design Level Metrics," Proc. 7th Working Conference on Reverse Engineering (WCRE'00), 23 - 25 Nov., 2000, pp. 248 - 256, Brisbane, Australia, 2000.

- [7] J.C.Cherniavsky and C.H.Smith, C.: On Weyukers Axioms for Software Complexity Measures. IEEE Transaction on Software Engineering, Vol. 17, 1991, No. 6, pp. 636–638.
- [8] Sanjay Mishra and Ibrahim Akman “applicability of Weyker’s properties on OO metrics: some Misunderstandings” ComSIS Vol. 5, No. 1, June 2008.
- [9] Ajay Rana ,Soumi Ghosh and S K Dubey, “Comparative Study of Factors that Affects Maintainability, International Journal on Computer Science and Engineering”, Vol 3 (12), December 2011.
- [10] N.E.Fenton *software metrics* A Rigorous Approach Newyork :Chapman & Hall ,1991.
- [11] M. Kiewkanya, N. Jindasawat, and P. Muenchaisri, “A Methodology for Constructing Maintainability Model of Object-Oriented Design,” Proc. 4th International Conference on Quality Software, 8 - 9 Sept., 2004, pp. 206 - 213. IEEE Computer Society, 2004.
- [12] S. W. A. Rizvi and R. A. Khan, “Maintainability Estimation Model for Object- Oriented Software in Design Phase (MEMOOD), 2010.
- [13] Li, W.: Another Metric Suite for Object-Oriented Programming. Journal of Systems and Software, Vol. 44, 1998, pp. 155–162
- [14] M. Genero, E. Manso, A. Visaggio, and M. Piattini, “Building Measure-Based Prediction Models for UML Class Diagram Maintainability,” Journal of Empirical Software Engineering, vol. 12, no. 5, pp. 517 -549, 2007.
- [15] Chidamber, S.R.—Kemerer, C.F.: A Metrics Suite for Object Oriented Design.IEEE Transactions on Software Engineering, Vol. 20, 1994, No. 6, pp. 476–493.
- [16] Sheldon, F.T.—Jerath, K.—Chung, H.: Metrics for Maintainability of Class In-heritance Hierarchies. Journal of Software Maintenance 14, 3 May 2002, pp. 147–160.
- [17] ISO/IEC 9126-1, Institute of Electrical and Electronics Engineers, Part1: Quality Model, 2001.
- [18] Gursaran, G.R.: On the Applicability of Weyuker Property Nine to Object-Oriented Structural Inheritance Complexity Metrics. IEEE Transaction on Software Engineering, Vol. 27, 2001, No. 4, pp. 361–364.
- [19] Abreu, F.B.—Carapuca, R.: Candidate Metrics for Object-Oriented Software within a Taxonomy Framework. Journal of System Software, Vol. 26, 1994,pp.87–96.
- [20] M. Genero, J. Olivas, M. Piattini, and F. Romero, “A Controlled Experiment for Corroborating the Usefulness of Class Diagram Metrics at the Early Phases of Object-Oriented Developments,” Proc. of the ADIS2001, Workshop on Decision Support in Software Engineering, vol. 84. Spain, 2001.
- [21] Henderson-Sellers, B.: Object Oriented Metrics: Measures of Complexity. Pren-tice Hall PTR: Englewood Cliffs, NJ, 1996; pp. 130–132.
- [22] Amjan Shaik,C. R. K. Reddy, Bala Manda,Prakashini. C,Deepthi. Metrics for Object Oriented Design Software Systems: A Survey , Journal of Emerging Trends in Engineering and Applied Sciences (JETEAS) 1 (2): 190-198.
- [23] H.Zuse “properties of software measures” software quality j..vol-1 pp225-260,1992.
- [24] Weyuker, E. J.: Evaluating Software Complexity Measures. IEEE Transactions on Software Engineering, Vol. 14, 1988, No. 9, pp. 1357–1365.
- [25] M.S.Ranwat,A.Mittal, S.K.Dubey Survey on impact of software metrics on software quality (IJACSA)International journal of Advanced Computer Science and Applications, Vol.3, No.1, 2012.

#### AUTHORS PROFILE



Computer Science & Engineering, S.S.N. Engg. College, Ongole, Andhra Pradesh.

**Mr. D.N.V. Syma Kumar** received B.Tech Degree from JNTU, Hyderabad and M.Tech from BHARAT University, Chennai. He is currently pursuing his Ph.D. from Department of Computer Science, Krishna University, Andhra Pradesh. Presently he is working as an Associate Professor in the Department of



Computer Science & Engineering, Acharya Nagarjuna University. His current research is focused on Software engineering. He has published several papers in National & International Journals.

**Dr. R. Satya Prasad** received Ph.D. Degree in computer science in the faculty of Engineering in 2007 from Acharya Nagarjuna University, Andhra Pradesh. He received gold medal from Acharya Nagarjuna University for his outstanding performance in master’s degree. He is currently working as Associate Professor

Appendix-A

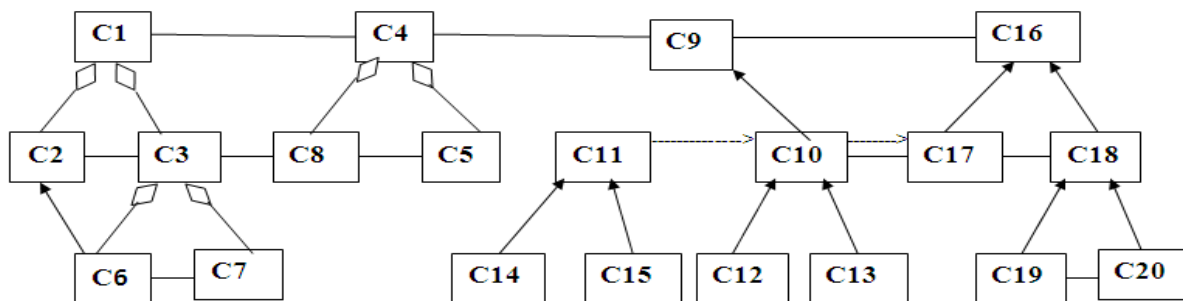


Figure-1 :UML Class representation base diagram Weyker's Properties validation.

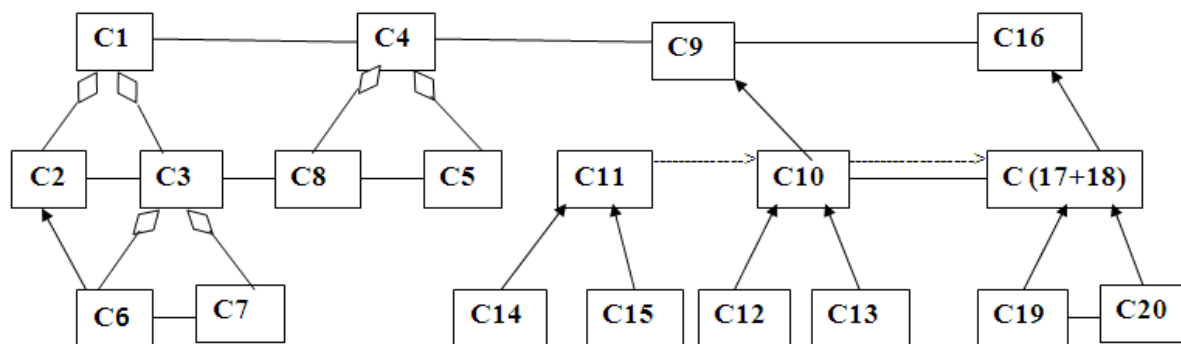


Figure-2: UML Class representation diagram Weyker's Property 5(i) validation.

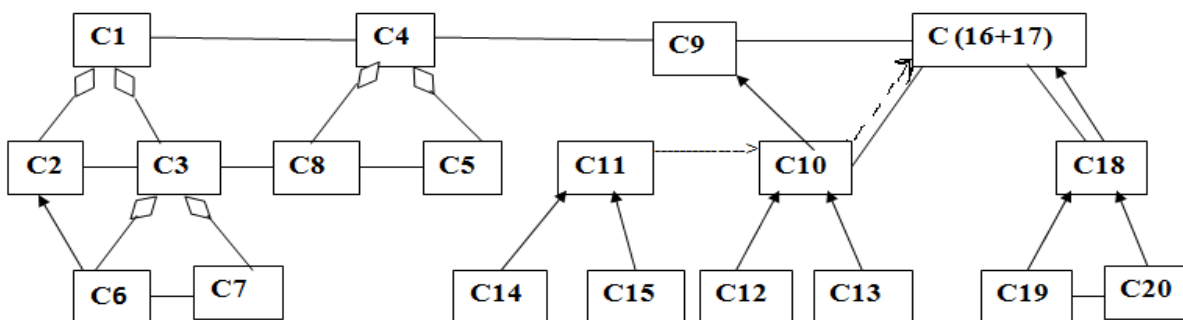


Figure-3: UML Class representation diagram Weyker's Property 5(ii) validation.



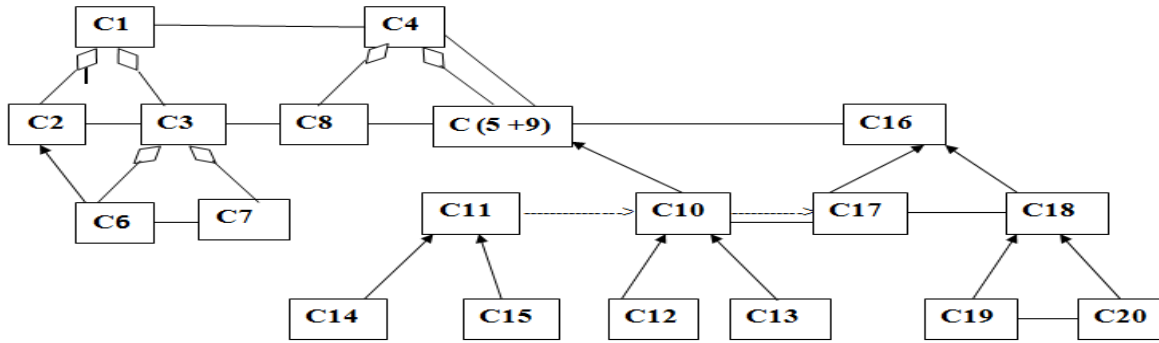


Figure-4: UML Class representation diagram Weyker's Property 5(iii) validation.

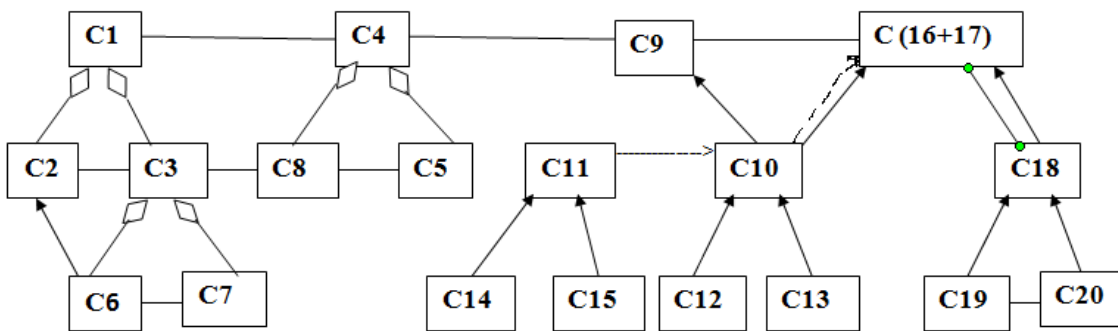


Figure-5: UML Class representation diagram Weyker's Property 6-a validation.

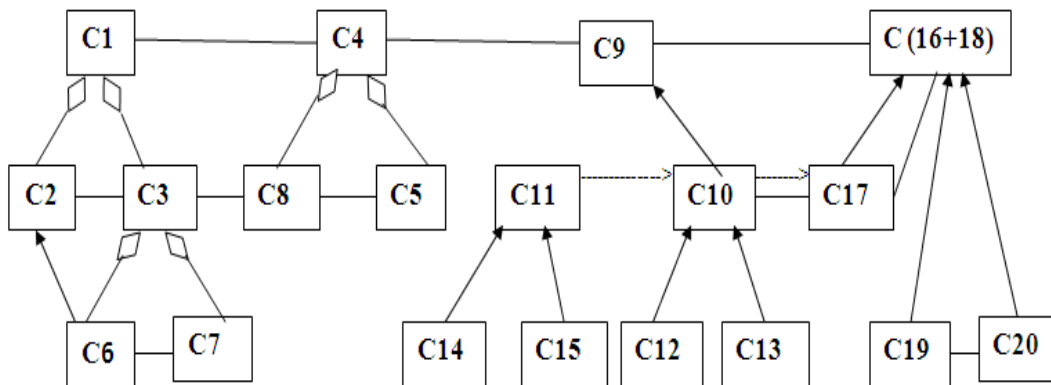


Figure-6: UML Class representation diagram Weyker's Property 6-b validation.

# A hybrid classification algorithm and its application on four real-world data sets

Lamiaa M. El Bakrawy <sup>#1</sup>, Abeer S. Desuky <sup>\*2</sup>

<sup>#</sup> Faculty of Science, Al-Azhar University  
Cairo, Egypt

<sup>1</sup> lamiaabak@yahoo.com.

<sup>\*</sup> Faculty of Science, Al-Azhar University  
Cairo, Egypt

<sup>2</sup> abeerdesuky@yahoo.com.

**Abstract**—The aim of this paper is to propose a hybrid classification algorithm based on particle swarm optimization (PSO) to enhance the generalization performance of the Adaptive Boosting (AdaBoost) algorithm. AdaBoost enhances any given machine learning algorithm performance by producing some weak classifiers which requires more time and memory and may not give the best classification accuracy. For this purpose, We proposed PSO as a post optimization procedure for the resulted weak classifiers and removes the redundant classifiers. The experiments were conducted on the basis of four real-world data sets: Ionosphere data set, Thoracic Surgery data set, Blood Transfusion Service Center data set (btsc) and Statlog (Australian Credit Approval) data set from the machine-learning repository of University of California. The experimental results show that a given boosted classifier with our post optimization based on particle swarm optimization improves the classification accuracy for all used data. Also, The experiments show that the proposed algorithm outperforms other techniques with best generalization.

## I. INTRODUCTION

Nowadays there is tremendous amount of data being collected and stored in databases everywhere across our realm. It is easy now to find databases with Terabytes - about 1,099,511,627,776 bytes - of data in enterprises and research fields. Numerous invaluable information and knowledge buried in such databases; and without facile methods for extracting this buried information it is practically impossible to mine for them. Many algorithms were created throughout the decades for extracting what is called nuggets of knowledge from large sets of data. There are several diverse methodologies to approach this problem: classification, clustering, association rule, etc. our paper will focus on classification [1] [2] [3].

Classification is one of the most frequently studied problems by data mining and machine learning researchers [2]. Classification consists of predicting a certain outcome based on a given input. A classifier is a function or an algorithm that maps every possible input (from a legal set of inputs) to a finite set of classes or categories[4]. Adaptive Boosting (AdaBoost) is a widespread successful technique used to boost the classification performance of weak learner. Hu et al. [5] proposed two algorithms based on AdaBoost classifier for online intrusion

detection. They used the traditional AdaBoost where decision stumps are used as weak classifiers in the first algorithm. In the second algorithm, online Gaussian mixture models (GMMs) are used as weak classifiers to improve online AdaBoost process. The second algorithm showed a better performance in the experiments than the traditional AdaBoost process that uses decision stumps. Another improved AdaBoost algorithm named (ISABoost) proposed by X. Qian et al. [6] and applied in scene categorization. In ISABoost the inner structure of each trained weak classifier is adjusted before the traditional weights determination process. ISABoost algorithm after inner structure adjusting in each iteration of AdaBoost learning selects an optimal weak classifier and determines its weight. Three scene data sets used in Comparisons of ISABoost and traditional AdaBoost algorithms, where Back-propagation networks and SVM are served as weak classifiers, and ISABoost verified its effectiveness.

Choi et al. [7] presented a novel multiple classifier system -termed "classifier ensemble"- based on AdaBoost for tackling false-positive (FP) reduction problem in Computer-aided Detection (CADe) systems, especially of mass abnormalities on Mammograms. Different feature representations were combined with data resampling based on AdaBoost learning to create the "classifier ensemble". Adjusting the size of a resampled set is the effective mechanism used by classifier ensemble to regulate the degree of weakness of the weak classifiers of conventional AdaBoost ensemble. Support Vector Machines (SVM) and Neural Network (NN) with back-propagation algorithm were used as base classifiers and applied on Digital Database for Screening Mammography (DDSM) DB. The area under the Receiver Operating Characteristics (ROC) was the used criterion to evaluate the classification performance and the comparative results showed the potential clinical effectiveness of the proposed ensemble.

As the AdaBoost approach produces a large number of weak classifiers, Particle Swarm Optimization (PSO) has the potential to automatically elect a good set of weak classifiers for AdaBoost and improve the algorithm performance. Our goal is to optimize the AdaBoost algorithm performance using the Particle Swarm Optimization technique.

The rest of this paper is organized as follows: Brief introduction of AdaBoost and Particle Swarm Optimization (PSO) algorithms are introduced in Section (II). The details of the proposed algorithm is presented in Section (III). Section (IV) shows the data sets and experimental results. Conclusions are discussed in Section (V).

### A. AdaBoost

The AdaBoost algorithm is frequently used boosting method proposed by Freund and Schapire [8] which was a kind of boosting algorithms proposed by Schapire in 1990 [9]. AdaBoost algorithm is used to boost the classification performance of a weak learning classifier by combining iteratively a collection of weak classifiers to form a stronger classifier. The weak classifiers are combined by AdaBoost by taking into account a weight distribution on the training samples such that more weight is assigned to samples misclassified by the previous iterations [10].

The AdaBoost algorithm takes as input a training set  $(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)$  where each instance  $x_i \in X$ , and each label  $y_i \in \{-1, +1\}$  [8][10][11].

On each iteration  $t = 1, \dots, T$ , a distribution  $D_t$  is computed over the  $m$  training instances where  $T$  is the number of weak classifiers  $c_t$ :

$$D_m^{(i)} = 1/m, i = 1, 2, \dots, m. \quad (1)$$

The given weak classifier or weak learning algorithm is applied in each iteration  $t$  to find a weak hypothesis

$$c_t : X \longrightarrow \{-1, +1\}, \quad (2)$$

Where the aim of the weak learner is to find a weak hypothesis with low weighted error  $\epsilon_t$  relative to  $D_t$

$$\epsilon_t = \sum_{i=1}^m D_t^{(i)} [y_i \neq c_t(x_i)]. \quad (3)$$

then update the Distribution  $D_t$  over the  $m$  training examples in each iteration

$$D_{t+1}^{(i)} = D_t^{(i)} \exp(-\alpha_t y_i c_t(x_i)) / Z_t, i = 1, 2, \dots, m. \quad (4)$$

where,

$$\alpha_t = \frac{1}{2} \ln \frac{1 - \epsilon_t}{\epsilon_t} \quad (5)$$

and  $Z_t$  is a normalization factor. After  $T$  iterations the final strong classifier is:

$$c(X) = \text{sign}(\sum_{t=1}^T \alpha_t c_t(X)) \quad (6)$$

### B. Particle Swarm Optimization

Particle Swarm Optimization (PSO) was proposed by Kennedy and Eberhart in [12]. PSO is one of the latest evolutionary optimization techniques for continuous optimization problems. It simulates the social behavior of bird flocks or fish schools. In this social group, there is a leader who presents the best performance and guides the movement of the whole

swarm. The movement of each particle is directed by the leader and its own knowledge. Thus, the behavior of each particle is a compromise between its individual memory and a collective memory [13], [14], [15].

The canonical PSO algorithm consists of a swarm of particles, which are initialized randomly with a population of candidate solutions. They move iteratively through the  $n$ -dimension problem space to search the new solutions, where the fitness function can be calculated as the certain qualities measure. Each particle  $i$  has a position represented by a position-vector  $\vec{x}_{ij}$  ( $i$  and  $j$  are the index of the particle and its dimension), and a velocity represented by a velocity-vector  $\vec{v}_{ij}$  [16], [17], [18]. In each iteration  $t$ , the update of the velocity from the previous velocity to the new velocity is calculated by Eq. (7) .

$$v_{ij}(t+1) = \begin{cases} wv_{ij}(t) + c_1r_1(P_{ij}(t) - x_{ij}(t)) \\ +c_2r_2(G_j(t) - x_{ij}(t)) \end{cases} \quad (7)$$

where  $w$  is called the inertia weight, which governs how much the pervious velocity should be retained from the previous time step ,  $r_1$  and  $r_2$  are random numbers chosen in the interval  $[0,1]$  for the  $j$ -th dimension of the  $i$ -th particle.  $c_1$  and  $c_2$  is are positives constants called as social and cognitive coefficients,  $P_{ij}(t)$  is the personal best (pbest) position, which is defined as a vector with the best fitness function value achieved by the particle in the past and  $G_j(t)$  is the global best (gbest) is used to record the position with the best fitness function value achieved by the all the particles in the past. The new position is then determined by the sum of the previous position and the new velocity by Eq. (8).

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1). \quad (8)$$

From Eq. (7), a particle decides where to move next, considering its own experience, which is the memory of its best past position, and the experience of its most successful particle in the swarm. In the particle swarm algorithm, the particle searches the solutions in the problem space with a range  $[-s, s]$ . In order to prevent the particle from flying away out of the search scope, the velocity is restricted on the interval  $[-v_{max}, v_{max}]$  given in Eq.(9):

$$v_{ij} = \text{sign}(v_{ij}) \min(|v_{ij}|, v_{max}). \quad (9)$$

Where the value of  $v_{max}$  is  $p \times s$ , with  $0.1 \leq p \leq 1.0$  and is usually chosen to be  $s$ , i.e.  $p = 1$ . A given maximum number of iterations, number of iterations without improvement, or minimum fitness function error can be used as a stop criterion.

## II. THE PROPOSED ALGORITHM

AdaBoost algorithm is a sequential forward search procedure based on the greedy selection strategy. Because of this strategy, the resulted weak classifiers and their coefficients are not optimal. We proposed PSO as a post optimization procedure for the resulted weak classifiers, removes the redundant

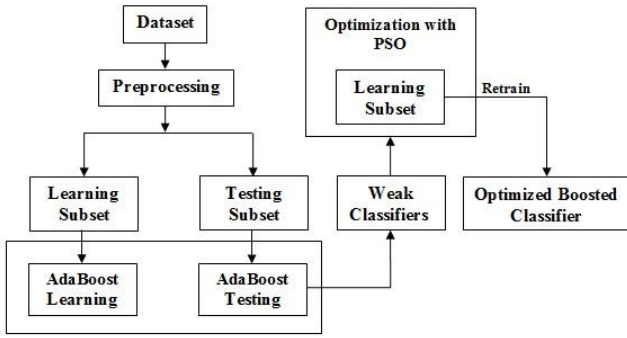


Fig. 1. The proposed algorithm

TABLE I  
PROPERTIES OF DATA SETS

Properties of data sets	Ionosphere	Thoracic Surgery	btsc	Statlog
No. Of classes	2	2	2	2
No. Of examples	351	470	748	690
No. Of attributes	34	17	5	14

TABLE II  
COMPARISON OF THE BOOSTED CLASSIFIERS WITHOUT AND WITH  
POST-OPTIMIZATION FOR IONOSPHERE DATA SET

AdaBoost classifiers	PSO_AdaBoost accuracy	AdaBoost accuracy
8	87.8 %	86.85 %
16	87.92 %	85.14 %
24	88.03 %	86.28 %
32	88.51 %	87.42 %
40	88.76 %	88.57 %
48	89.25 %	88.57 %
Average	88.37 %	87.13 %

classifiers and leads to shorter and better final classification performance.

According to the model of the AdaBoost algorithm that the weighted combination of weak learning classifiers  $\{c_1, c_2, \dots, c_T\}$  compose  $c(X) = \text{sign}(\sum_{t=1}^T \alpha_t c_t(X))$ , the final boosted classifier. Unfortunately the strong classifier comprises more weak classifiers requiring more memory and more time to evaluate and may produce less classification accuracy. To solve this, we propose a hybrid approach based on PSO as optimization algorithm. This algorithm (as shown in Fig. 1) elects the optimal weak classifiers with their weights as follows:

1<sup>st</sup> : Train the AdaBoost using the training data set. After T iterations of AdaBoost training, now we have T number of weak classifiers  $c_t$  with their weights  $\alpha_t$  which form the final strong classifier  $c(X)$ , then the performance of the system with the testing data set is measured.

2<sup>nd</sup> : Initialize particle population, each particle in the evolving population is a binary vector  $q$  composed of  $(q_1, q_2, \dots, q_T)$  denoting the weak classifiers which constitute the final strong classifier. The particles move iteratively through the X-dimension, where each particle has a position set to 1 or 0—denoting the weak classifier appearance or absence—which can be calculated according to the fitness function:

$$f(q) = 1 - E_q \quad (10)$$

where  $E_q$  is the error corresponding to the particle  $q$ . It is obvious that  $E_q$  expresses the fitness of  $q$  in the way that the smaller  $E_q$  is, the better  $q$  is. Each particle is updated in each iteration following the equations (7) – (9).

3<sup>rd</sup> : In this step we use the resulted best binary vector  $Q$  to determine the final classifiers with their corresponding weights and then the optimal boosted classifier is calculated.

### III. DATA SETS AND EXPERIMENTAL RESULTS

We have run our experiments using Matlab 12, on a system with a 2.30 GHZ Intel(R) Core(TM)i5 processor and 512 MB of RAM running Microsoft Windows 7 Professional (SP2).

The real-world data sets used throughout the paper to test our algorithm are: Ionosphere data set, Thoracic Surgery data

set, Blood Transfusion Service Center data set (btsc) and Statlog (Australian Credit Approval) data set. Data sets are obtained from the machine-learning repository of University of California [19]. A detailed description of the data sets is shown in Table I.

The parameters used for evolution were: In PSO,  $c_1 = c_2 = 2.0$ ,  $v_{max} = 4$ ,  $w$  was linearly decreased from 0.9 to 0.4, The PSO terminates if there is no found better individual within the next 10 generations.

From the experiment results shown in Table II, we can see that the accuracy of the classification for Ionosphere data set is increased by 1.24 % due to the Particle Swarm Optimization. The accuracy of the classification for Thoracic Surgery data set and Blood Transfusion Service Center data set are also increased by 1.57 % and 2.81 % respectively, as shown in Tables III and IV. While Table V shows that accuracy of the classification for Statlog data set is slightly increased by 0.32 %.

The results along with the comparison to other existing methods using Ionosphere data set, Thoracic Surgery data set, Blood Transfusion Service Center data set and Statlog data set are shown in Tables VI, VII, VIII and IX respectively. We found that generally the results achieved by the proposed method are higher than other techniques for all used data sets. The experimental results have demonstrated that a given boosted classifier with our post optimization based on Particle Swarm Optimization increases the classification accuracy.

### IV. CONCLUSION

The main aim of this paper is presenting a hybrid algorithm based on PSO to enhance the classification performance of the final classifier resulted from AdaBoost algorithm. Through the implementation of our algorithm, the experiments were conducted on the basis of four real-world data sets from the machine-learning repository of University of California. The experimental results showed that a given boosted classifier

**TABLE III**  
COMPARISON OF THE BOOSTED CLASSIFIERS WITHOUT AND WITH  
POST-OPTIMIZATION FOR THORACIC SURGERY DATA SET

AdaBoost classifiers	PSO_AdaBoost accuracy	AdaBoost accuracy
8	83.6 %	82.13 %
16	81.87 %	81.28 %
24	80.68 %	80.43 %
32	82.33 %	80 %
40	82.11 %	80.43 %
48	81.77 %	78.72 %
Average	82.06 %	80.49 %

**TABLE IV**  
COMPARISON OF THE BOOSTED CLASSIFIERS WITHOUT AND WITH  
POST-OPTIMIZATION FOR BLOOD TRANSFUSION SERVICE CENTER DATA  
SET

AdaBoost classifiers	PSO_AdaBoost accuracy	AdaBoost accuracy
8	78.38 %	78.34 %
16	77.69 %	71.39 %
24	78.03 %	73.52 %
32	77.81 %	74.86 %
40	76.62 %	74.86 %
48	76.72 %	75.40 %
Average	77.54 %	74.73 %

**TABLE V**  
COMPARISON OF THE BOOSTED CLASSIFIERS WITHOUT AND WITH  
POST-OPTIMIZATION FOR STATLOG DATA SET

AdaBoost classifiers	PSO_AdaBoost accuracy	AdaBoost accuracy
8	83.94 %	83.76 %
16	83.82 %	83.47 %
24	84.46 %	83.76 %
32	83.76 %	84.92 %
40	85.97 %	84.92 %
48	86.02 %	85.21 %
Average	84.66 %	84.34 %

**TABLE VI**  
CLASSIFICATION ACCURACY COMPARISON OF MACHINE LEARNING  
ALGORITHMS USING IONOSPHERE DATA SET

classifiers	Classification Accuracy
BP Neural Network Classifier [20]	81.75 %
RBF Neural Network [20]	82.78 %
SVM Classifier [20]	84.46 %
PSO_SVM [20]	82.78 %
K-nn [21]	86.3%
Naive Bayes [21]	87.8 %
PSO_AdaBoost(proposed)	88.37 %

**TABLE VII**  
CLASSIFICATION ACCURACY COMPARISON OF MACHINE LEARNING  
ALGORITHMS USING THORACIC SURGERY DATA SET

classifiers	Classification Accuracy
PART[22]	76.59 %
Nave Bayes[23]	77.74 %
Multilayer Perceptron[23]	80.91 %
Boosted Nave Bayes[23]	78.32 %
Boosted Multilayer Perceptron[23]	80.70 %
Boosted J48[23]	79.34 %
PSO_AdaBoost(proposed)	82.06 %

**TABLE VIII**  
CLASSIFICATION ACCURACY COMPARISON OF MACHINE LEARNING  
ALGORITHMS USING BLOOD TRANSFUSION SERVICE CENTER DATA SET

classifiers	Classification Accuracy
Online discretization [24]	75.63 %
CAIM [24]	75.63 %
Modified CAIM without merging stage [24]	75.63 %
Modified CAIM with merging [24]	75.63 %
LR [25]	77.14 %
NN [25]	75.55 %
ELM[25]	76.20 %
GBM [25]	76.34 %
RF [25]	75.05 %
PSO_AdaBoost(proposed)	77.54 %

**TABLE IX**  
CLASSIFICATION ACCURACY COMPARISON OF MACHINE LEARNING  
ALGORITHMS USING STATLOG DATA SET

classifiers	Classification Accuracy
IDE3[26]	71.5 %
C4.5 [26]	84.2 %
AdaboostC4.5 [27]	84.01 %
Bagging NB [27]	77.81 %
Adaboost NB [27]	81.16 %
MAdaBoost SVM [28]	81.20 %
Single SVM[28]	79.86 %
Arc-x4 SVM [28]	78.94 %
AdaBoost SVM [28]	79.16 %
K-nn [21]	57.5 %
CN2 [21]	84.2 %
PSO_AdaBoost (proposed)	84.66 %

with our post optimization based on particle swarm optimization performed quite well and improved the classification accuracy for all four data sets used with maximum accuracy increasing 2.81 % for Blood Transfusion Service Center data and minimum accuracy increasing 0.32 % for Statlog data set. The experiments also showed that the proposed algorithm outperforms other techniques applied on the same data.

## REFERENCES

- [1] Komal Arunjeet Kaur, Shelza Garg, "International Journal for Science and Emerging Technologies with Latest Trends", Vol. 17, No. 1, pp. 9-13, 2014.
- [2] Ravi Sanakal, Smt. T Jayakumari, "Prognosis of Diabetes Using Data mining Approach-Fuzzy C Means Clustering and Support Vector Machine", International Journal of Computer Trends and Technology (IJCTT), Vol. 11, No. 2, May 2014.
- [3] Ian H.Witten and Eibe Frank, "Data Mining: Practical Machine Learning Tools and Techniques", Second Edition, Morgan Kaufmann Publishers, Elsevier Inc. 2005.
- [4] Jayshri D. Dhande and D.R. Dandekar, "PSO Based SVM as an Optimal Classifier for Classification of Radar Returns from Ionosphere", International Journal on Emerging Technologies, Vol. 2, No. 2, pp. 1-3, 2011.
- [5] Hu W., Gao J., Wang Y., Wu O., and Maybank S.J., "Online Adaboost-Based Parameterized Methods for Dynamic Distributed Network Intrusion Detection", IEEE T. Cybernetics, pp.66-82, 2014.
- [6] Qian X., Tang Y. Y., Yan Z., Hang K., "ISABOOST: A weak classifier inner structure adjusting based Adaboost algorithm-ISABOOST based application in scene categorization", Neurocomputing 103 Published by Elsevier, pp. 104-113, 2013.
- [7] Choi, J. Y., Kim, D. H., Plataniotis, K. N., and Ro, Y. M., "Combining Multiple Feature Representations and AdaBoost Ensemble Learning for Reducing False-Positive Detections in Computer-aided Detection of Masses on Mammograms", 34th Annual International Conference of the IEEE EMBS San Diego, California USA, 2012.

- [8] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm", in Proceedings of the 13th International Conference on Machine Learning, Bari, Italy, pp. 148-156, 1996.
- [9] R.E. Schapire, "The Strength of Weak Learnability", Mach. Learn., Vol. 5, No. 2, pp. 197-227, 1990.
- [10] Zhengjun Cheng, Yuntao Zhang, Changhong Zhou, Wenjun Zhang, Shibo Gao, "Classification of Skin Sensitizers on the Basis of Their Effective Concentration 3 Values by Using Adaptive Boosting Method", International Journal of Digital Content Technology and its Applications (JDCTA), Vol. 4, No. 2, pp. 109 - 121, 2010.
- [11] Jianfang Cao, Junjie Chen, and Haifang Li, "An Adaboost-Backpropagation Neural Network for Automated Image Sentiment Classification", The Scientific World Journal, vol. 2014, 2014. doi:10.1155/2014/364649
- [12] Kennedy J, Eberhart R: Particle swarm optimization. In Proceedings International Conference on Neural Networks (ICNN 95) Perth, Australia; 1942-1948.
- [13] Wu J., "Integrated Real-Coded Genetic Algorithm and Particle Swarm Optimization for Solving Constrained Global Optimization Problems", Advances in Information Technology and Education Communications in Computer and Information Science, Vol.201, pp. 511-522, 2011.
- [14] Qiu X., Lau H., "An AIS-based hybrid algorithm for static job shop scheduling problem", Journal of Intelligent Manufacturing, Vol. 25, Issue 3, pp. 489-503, 2014.
- [15] Hasan S., Shamsuddin S., Yusob B., "Enhanced Self Organizing Map (SOM) and Particle Swarm Optimization (PSO) for Classification", Jurnal Generic, Vol. 5, pp. 7-11, 2010.
- [16] Soliman M., Hassanien A., Ghali N., Onsi H., "An adaptive Watermarking Approach for Medical Imaging Using Swarm Intelligent", International Journal of Smart Home, Vol. 6, No. 1, pp. 37-50, January 2012.
- [17] Costa Jr S., Nadia Nedjah N., Mourelle L., Automatic Adaptive Modeling of Fuzzy Systems Using Particle Swarm Optimization, Transactions on Computational Science VIII, Lecture Notes in Computer Science, Vol. 6260, pp. 71-84, 2010.
- [18] Sun J., Palade V., Cai Y., Fang W., Wu X., Biochemical systems identification by a random drift particle swarm optimization approach, BMC Bioinformatics 15(Suppl 6):S1 <http://www.biomedcentral.com/1471-2105/15/S6/S1>, pp. 1-17, 2014.
- [19] <http://archive.ics.uci.edu/ml/datasets.html>
- [20] Dhande M., Dandekar D., Badjate S., Performance Improvement of ANN Classifiers using Pso, National Conference on Innovative Paradigms in Engineering & Technology (NCIPET\_2012), Proceedings published by International Journal of Computer Applications (IJCA), pp. 32-36, 2012.
- [21] Hacibeyoglu M., Arslan A., Kahraman S., Improving Classification Accuracy with Discretization on Datasets Including Continuous Valued Features, International Scholarly and Scientific Research & Innovation 5(6), 2011.
- [22] Sindhu V., Prabha S., Veni S., Hemalatha M., THORACIC SURGERY ANALYSIS USING DATA MINING TECHNIQUES, Sindhu V et al, Int.J.Computer Technology & Applications, Vol 5 (2), pp. 578-586, 2014.
- [23] Harun A., Alam N., Predicting Outcome of Thoracic Surgery by Data Mining Techniques, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 1, January 2015.
- [24] Vora S., Mehta R., MCAIM: Modified CAIM Discretization Algorithm for Classification, International Journal of Applied Information Systems (IJAIS) ISSN : 2249-0868, Foundation of Computer Science FCS, New York, USA, Volume 3 No.5, July 2012
- [25] Wang Y., Li Y., Xiong M., Jin L., Random Bits Regression: a Strong General Predictor for Big Data, eprint arXiv:1501.02990, Jan 2015.
- [26] Anyanwu M., Shiva S., Comparative Analysis of Serial Decision Tree Classification Algorithms, International Journal of Computer Science and Security, (IJCSS) Volume (3) : Issue (3), 2009.
- [27] Ferdousy E., Islam M., Matin A., Combination of Naive Bayes Classifier and K-Nearest Neighbor (cNK) in the Classification Based Predictive Models, Computer and Information Science; Vol. 6, No. 3, 2013.
- [28] Wang, Shi-jin and Mathew, Avin D. and Chen, Yan and Xi, Li-feng and Ma, Lin and Lee, Empirical analysis of support vector machine ensemble classifiers. Expert Systems with Applications, 36(3, Part 2). pp. 6466-6476, 2009.

# Towards an Intelligent Decision Support System Based on the Multicriteria K-means Algorithm.

Dadda Afaf

Department of Industrial and Production Engineering,  
ENSAM University My ISMAIL  
Meknes, Morocco.

Brahim Ouhbi

Department of Industrial and Production Engineering,  
ENSAM University My ISMAIL  
Meknes, Morocco.

**Abstract**—the actual management of Renewable Energy (RE) project is involving a large number of stakeholders in an uncertainty and dynamic environment. It is also a multi-dimensional process, since it has to consider technological, financial, environmental, and social factors. Multicriteria Decision Analysis appears to be the most appropriate approach to understand the different perspectives and to support the evaluation of RE project.

This paper aims to present an intelligent decision support system (IDSS), applied to renewable energy field. The proposed IDSS is based on combination of the binary preference relations and the multi-criteria k-means algorithm. An experimental study on a real case is conducted. This illustrative example demonstrates the effectiveness and feasibility of the proposed IDSS

**Keywords**—Artificial Intelligence; Decision Support system, Multicriteria relation Clustering; k-means algorithm.

## I. INTRODUCTION

The last decade have known a new deploy of Intelligent Decision Support System (IDSS). In the case of Renewable Energy (RE), the deal has just begun and many researchers have tried to build and improved new DSS enable to support and assist their local decision makers [1]. Moreover, the RE field provides considerable new challenges given that it requires more intelligent algorithms and models that can solve problems in an uncertainty and dynamic environment. It is widely approved that the real decision-making process is multi-dimensional, including numerous stakeholders and considering various factors. Therefore, Multi-criteria Decision Analysis (MCDA) is the most appropriate approach to understand the different perspectives and to support the development of appropriate intelligent algorithm. This research aims to improve a local IDSS able to assist stakeholders on their real decision problems. For this purpose a new relational multi-criteria model is proposed based on the multi-criteria k-means algorithm defined by De Smet and al. [8] and on the preference function commonly used on the PROMETHEE method.

This paper is organized as follows: the first section presents the related works. The second section explains the basic concepts to be used in the improvement of our model. The third section

presents the proposed IDSS. Finally, an illustrative example is given in section five. This example is applied to solve local renewable energy project management.

## II. RELATED WORKS

The IDSS approach is combining the Decision Support System (DSS) and the Artificial Intelligent (AI) techniques.

The DSS is a tool based information system that support decision makers in various levels such as planning, managing and organizing. In practice, it used databases and models to solve complex problems. Recent research demonstrated that the introduction of the AI techniques, enable the decision support system to reproduce the human capabilities as closely possible. Some research in IDSS [2], focused on enabling systems to encode the cognitive behaviors of human experts using predicate logic rules. So, an ideal IDSS [3] is able to work like a human consultant who is supporting decision making by analyzing, identifying, diagnosing problems and finally proposing possible solutions.

Multicriteria Decision Analysis (MCDA) deals with the process of making decisions in the presence of multiple objectives [4]. This approach can be used to solve three main problems: the choice problem, the ranking problem and the sorting problems. In MCDA field, different methods that are based on the following approaches can be used:

- The top-down approach, seeks to aggregate the “n” criteria into a single criterion, it supposed that the judgment are transitive (ex:  $a > b$  and  $b > c$  so  $a > c$ ). Example of such approach is the AHP/ANP and the MAUT method.
- The Bottom-up approach, tries to compare potential alternatives to each other and set up relationships between themes. PROMETHEE and The ELECTRE method are examples of such approach.
- The local aggregation, which tries to find an ideal solution in the first step. Then, proceeds to an iterative search to find a better solution. Such as the VIKOR and TOPSIS method.

In the multicriteria decision aid field, a lot of attention has been given to the relational multicriteria clustering. The clustering is

the process that categorizes data into clusters. Each cluster includes elements, which have a maximum similarity with each other and maximum dissimilarity with the element of other clusters [5]. Clustering is mainly studied in data mining field, where large data sets with many attributes of different types are considered. De Smet and al. [6] have proposed a new extension of the traditional k-means algorithm. Its contribution is focused on the use of the binary preference relations between the actions to define the distance between clusters. This new concept of multicriteria distance; make the use of the k-means algorithm in the multicriteria decision making context more feasible. However, the real utilization of this new multicriteria algorithm is a difficult task. In consequence, this study tries to propose the use of the relational clustering approach in the determination of the new profile of each centroid. The next section will present the basic concepts that were used in the conception of the proposed hybrid algorithm.

### III. RELATIONAL MULTICRITERIA DECISION AID CLUSTERING

#### A. Basic Concepts

Let  $A = \{a_1, a_2, \dots, a_n\}$  be a set of “n” elements called ‘alternatives’ and let  $C = \{c_1, c_2, \dots, c_m\}$  be a set of “m” elements called ‘criteria’. The clustering model considers the following relations: Preference (P), Indifference (I), and Incomparability (J), which result in the comparison between two actions  $a_i$  and  $a_j$

$$\left\{ \begin{array}{l} a_i P a_j \text{ if } a_i \text{ is preferred to } a_j \\ a_i I a_j \text{ if } a_i \text{ is indifferent to } a_j \\ a_i J a_j \text{ if } a_i \text{ is incomparable to } a_j \end{array} \right\} \quad (1)$$

The three relations  $\langle P; I; J \rangle$  make up a preference structure on ‘A’. If both elements  $a_i$  and  $a_j$  are compared, one and only one of the following properties are true:  $a_i P a_j$  or  $a_i I a_j$  or  $a_i J a_j$ .

The criteria independency or dependency is an important question that needs to be analyzed, because the clustering procedure takes into account the preference scale implied by the criteria. Therefore, in this research, the criteria set,  $C = \{c_1, c_2, \dots, c_m\}$  is split into the selectability criteria and rejectability criteria. For this purpose, De Smet and al. [8] were interested on the extension of the k-means algorithm to the multicriteria structure. The basic cognitive process behind the method is that all alternatives within the same cluster are preferred, indifferent and incomparable. Due to the multicriteria nature of the problem, the concept of Euclidean distance, so widely used in classification techniques, does not appear to be appropriate. To avoid this limitation, a multicriteria distance based on the preference structure is defined as:

$$d(a_i, a_j) = 1 - \frac{\sum_{k=1}^4 |P_k(a_i) \cap P_k(a_j)|}{n} \quad (2)$$

$P_{k=1,2,3,4}$  represents the profile of each alternatives  $a_i$   $\{P_1=I(a_i), P_2=P^+(a_i), P_3=P^-(a_i), P_4=J(a_i)\}$ .

In the extended version of k-means algorithm, De Smet and al., the choice of the centroids of the cluster is based on a voting procedure. In their published article, Broudi and Bahloul [14], had observed that the use of relation between clusters is based on the detection of the dominant relation between the alternatives belonging to the same different clusters.

Formally, the frequency of occurrence of a preference relation when considering a couple of clusters  $C_u$  and  $C_v$  is as follows:

$$Q_{Suv} = \left| \left\{ (a_i, a_j) \in A \mid a_j \in C_u \Delta a_j \in C_v \Delta a_i S a_j \right\} \right| \forall S \in \{I, R, P^+, P^-\}. \quad (3)$$

The proposed measure  $Q_{Suv}$  is computed for each preference relation  $S$  belonging to the set  $\{I, R, P^+, P^-\}$ . Then the dominant relation can be determined by

$$S = \arg \max (Q_{Suv}) \quad (4)$$

### IV. THE PROPOSED MODEL

#### A. Proposed Algorithm

The proposed algorithm offers the possibility to order the resulting clustering, which makes an ordered multicriteria clustering (OMCC) (see Figure I).

The idea behind the improvement of this multicriteria algorithm is based on the introduction of two concepts [7]:

- The use of the ideal and negative ideal solution as initial centroids:

$$A^+ = (\max a_{ij}^n / j \in J) \text{ and } A^- = (\min a_{ij}^n / j \in J'), \quad (5)$$

where “J” is the number of select ability criteria and J’ is the number of reject ability criteria.

- The second main contribution of this work is due to the introduction, for each category of criteria, of a select-ability function  $f'_s$  and a reject-ability function  $f'_r$  defined as:

$$f_s(A_i) = \frac{\sum_{j=1..J} W_j^S a_j^n(a_i)}{\sum_{i=1..n} \sum_{j=1..J} W_j^S a_j^n(a_i)} \quad (6)$$

$$f_r(A_i) = \frac{\sum_{j=1..J} W_j^R a_j^n(a_i)}{\sum_{i=1..n} \sum_{j=1..J} W_j^R a_j^n(a_i)} \quad (7)$$

The index “j” represents the number of the select-able or the reject-able criteria.

The proposed algorithm is an extension of the multicriteria k-means algorithm presented by De Smet and al. [13].

**Step 0:** The starting point of this approach is a decision matrix, which had the following structure:

$$[c_1 \dots c_m]$$



$$\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix}$$

Where:

$A = \{a_1, a_2, \dots, a_n\}$  is a set of the evaluated alternatives,  
 $C = \{c_1, c_2, \dots, c_m\}$  is a set of criteria according to which the decision problem will be evaluated,  
 $a_{ij}$  is the value of an alternative  $i$  to a criterion  $j$ .

**Step 1:** Splitting the criteria into two categories: Select-ability criteria (Set of criterion to Maximize/Benefit) and Reject-ability ones (Set of criterion to Minimize/Cost). Note that each criterion  $c_i$  had a weight  $w_i$  that reflects the decision maker preferences.

The vector of weightings  $W = \{w_1, w_2, \dots, w_m\}$  should respect the following conditions:

$$\forall w_i \in W w_i \geq 0 \quad \sum_{i \in 1, \dots, m} w_i = 1$$

**Step 2:** Calculate a normalized version of the initial decision matrix; it's a widely used normalization formula [15]. This step is very important in order to unify the different unit of each criterion. The structure of the new matrix can be expressed as follows:

$$a_{ij}^n = \frac{a_{ij} - a_j^{\min}}{a_j^{\max} - a_j^{\min}} \quad i = 1, \dots, n; j \in \Omega_b \quad (8)$$

$$a_{ij}^n = \frac{a_j^{\max} - a_{ij}}{a_j^{\max} - a_j^{\min}} \quad i = 1, \dots, n; j \in \Omega_c \quad (9)$$

where:

$$a_j^{\min} = \min_{1 \leq i \leq n} \{a_{ij}\}, \quad a_j^{\max} = \max_{1 \leq i \leq n} \{a_{ij}\} \quad (10)$$

$\Omega_b$  and  $\Omega_c$  are respectively the index sets of benefit criteria and cost criteria and  $a_{ij}^n$  is the normalized value of  $a_{ij}$ .

**Step3:** Determine the ideal and negative-ideal solutions:

$$A^+ = (\max a_{ij}^n / j \in J) \text{ and } A^- = (\min a_{ij}^n / j \in J') \quad (11)$$

**Step4:** Calculate the select ability function  $f_s$  and reject ability function  $f_r$  as defined earlier.

Afterward the profile  $P(a_i)$  of each alternative is determined

$$\{P_1=I(a_i), P_2=P^+(a_i), P_3=P^-(a_i), P_4=J(a_i)\}$$

$$P^+(a_i) = \left\{ \begin{array}{l} a_j \in A \setminus \{a_i\} \text{ and } f_s(a_i) - f_s(a_j) < 0 \\ \text{and } f_r(a_i) - f_r(a_j) > 0 \end{array} \right\} = P_1(a_i) \quad (12)$$

$$P^-(a_i) = \left\{ \begin{array}{l} a_j \in A \setminus \{a_i\} \text{ and } f_s(a_i) - f_s(a_j) > 0 \\ \text{and } f_r(a_i) - f_r(a_j) < 0 \end{array} \right\} = P_2(a_i) \quad (13)$$

$$J(a_i) = \left\{ \begin{array}{l} a_j \in A \text{ and } f_s(a_i) - f_s(a_j) < 0 \\ \text{and } f_r(a_i) - f_r(a_j) < 0 \\ \text{Or } f_s(a_i) - f_s(a_j) > 0 \\ \text{and } f_r(a_i) - f_r(a_j) > 0 \end{array} \right\} = P_3(a_i) \quad (14)$$

$$I(a_i) = \left\{ \begin{array}{l} a_j \in A \text{ and } i \neq j \\ a_j \notin \{P^+(a_i) \cup P^-(a_i) \cup J(a_i)\} \end{array} \right\} = P_4(a_i) \quad (15)$$

**Step5:** Before starting the multicriteria k-means algorithm, the initialization of the centers by using the ideal and negative ideal solution will be done and in each, iteration the following calculate will be done.

- **In each Iteration** the distance between the profiles  $P(A_i)$  is calculate as follow:

$$d(a_i a_j) = 1 - \frac{\sum_{k=1}^4 |P_k(a_i) \cap P_k(a_j)|}{n} \quad (16)$$

- **In each Iteration** the profile  $P(r_i)$  of the centers of the new clusters  $C_i$  will be calculate as follow:

$$I(r_1) = \text{Argmax}_{j \in 1 \dots n} (A_j \text{ for } A_j \in \cup_{i, A_i \in C_1} I(A_i)) \quad (17)$$

$$P^+(r_1) = \text{Argmax}_{j \in 1 \dots n} (A_j \text{ for } A_j \in \cup_{i, A_i \in C_1} P^+(A_i)) \quad (18)$$

$$P^-(r_1) = \text{Argmax}_{j \in 1 \dots n} (A_j \text{ for } A_j \in \cup_{i, A_i \in C_1} P^-(A_i)) \quad (19)$$

$$J(r_1) = \text{Argmax}_{j \in 1 \dots n} (A_j \text{ for } A_j \in \cup_{i, A_i \in C_1} J(A_i)) \quad (20)$$

Before the applying this model, we need to specify the number of clusters. In this study, we suggest that the initial centers will include at the minimum the ideal alternative and the negative ideal alternative. Afterward, the alternatives are assigned to the nearest cluster. To achieve this operation the multicriteria distance based on the preference structure is used.

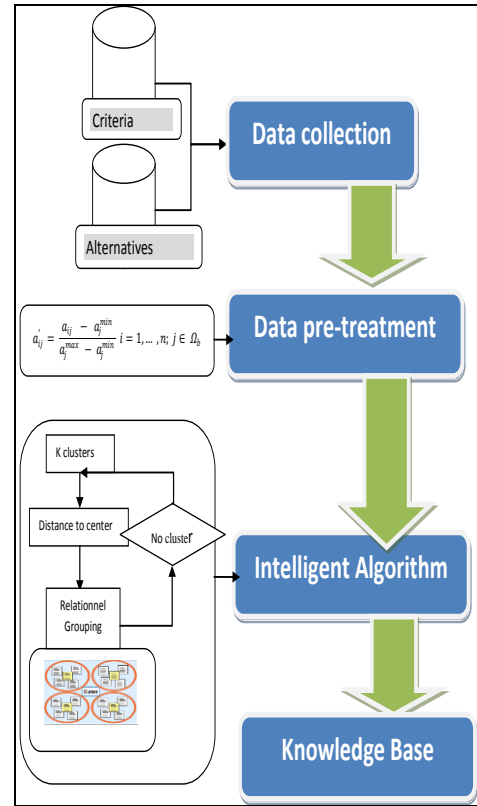


Figure 1 : IDSS Architecture

**B. The proposed IDSS**

The proposed IDSS is a cooperative DSS, which allows the decision maker to complete and refine the decision suggestions before sending them back to the system for validation. This system includes three fundamental components; the knowledge base, the logical model which is based on the above relational multicriteria algorithm and finally the user interface.

The early framework of the proposed IDSS consists of four phases: The first phase is the **data collection** it includes the criteria and the alternatives collection. The second important phase is the **data pretreatment** by determining the criteria dependency or independency, by calculating their weight and finally by determining the normalized decision matrix. The third phase is the **algorithm implementation**. The fourth phase is the **results and graphs generation**.

The main strengths of the proposed IDSS are:

- a. Simply use; the user have just to set up the alternatives; and the IDSS would propose a list of most important criteria that could be chosen by the user to evaluate the alternatives.
- b. The real-time decision-making and an interactive mode.
- c. A detailed database that includes specific information required to conduct the analysis of the various problems.
- d. A powerful model, in which the management system is responsible for the treatment, including the storage, the update and the adjustment.

**V. ILLUSTRATIVE EXAMPLE**

In this section, we consider the application of the presented approach to a real life problem which is the renewable energy project selection. Nowadays, many investors are interesting on implementing new renewable energy project around the world. The success of the decision making process regarding the selection of these projects, depends a lot on the effectiveness of the feasibility stage.

This problem has already been treated in the literature by [8, 9, 10 and 11]. In this research it was chosen to study six alternatives  $\{A_0, \dots, A_5\}$  according to a set of five criteria  $\{c_0, \dots, c_9\}$ . Technical, technological, financial and social criteria are considered in the selection of the optimal project (see Figure 2). Note that, only the criterion  $c_3$  is a benefit one, while all the others one is cost criterion. The example is based on a set of 6 projects, namely:  $\{A_0, A_1, A_2, A_3, A_4, A_5\}$ .

TABLE I DECISION MATRIX

	$c_0$	$c_1$	$c_2$	$c_3$	$c_4$	$c_5$	$c_6$	$c_7$	$c_8$	$c_9$
$A_0$	8	4	1	10	100	1	1	3	1	3

$A_1$	4	3	9	5	120	9	3	9	3	9
$A_2$	1	4	8	25	18	8	6	6	4	6
$A_3$	3	9	1	100	30	4	8	8	9	8
$A_4$	8	1	1	20	40	3	9	4	1	4
$A_5$	1	9	6	6	50	6	1	3	9	3

To illustrate how the proposed IDSS works, a simulation of different iterations from the algorithm process applied on a decision matrix is presented below. First of all, a preliminary phase consists of preparing the input for the algorithm (normalized decision matrix, calculate preference functions and finally identify sets of profiles for each alternative).

TABLE II PREFERENCE MATRIX

	$A_0$	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$
$A_0$	I	P-	P+	P-	P-	P-
$A_1$	P+	I	P+	P+	P+	P+
$A_2$	P-	P-	I	P-	P-	P-
$A_3$	P+	P-	P+	I	P+	P+
$A_4$	P+	P-	P+	P-	I	P-
$A_5$	P+	P-	P+	P-	P+	I

The result of this phase can be summarized in a preferences matrix (Table II). The created profiles will be the input of our algorithm. In our case it's a multicriteria k-means clustering algorithm based on the definition of the aforementioned distance between two alternatives.

An initial set of  $K=2$  clusters is generated first. We have used an approach of ideal and negative ideal solution in the constructing of this initial partition. The centroids are then generated from this initial set of clusters, through their profiles; In this case, while choosing 2 clusters on this input data, the algorithm proceeds on 3 iterations until the stabilization.

The relations between the created clusters could be explained as:  $C_i P^+ C_j$  means that all alternatives in  $C_j$  are better than alternatives in  $C_i$ , same for the relation  $P^-$ , a cluster  $C_i P^- C_j$  means that all alternatives in  $C_j$  are worse than alternatives in  $C_i$ . In addition, if  $K = 6$  (number of clusters equal number of alternatives), we can order the set of alternatives, in a condition that there are no incomparability relations between the different clusters. According to the above results, it can be

concluded that the  $A_3$  project is the best one, while projects  $A_0$ ,  $A_2$  and  $A_4$  are the worst ones.

## VI. CONCLUSION

This paper had presented an IDSS based on new relational multicriteria k-means algorithm, this system aimed to assist decision maker in their decision making process. The originality of this approach is due to the use of the research resulting from the clustering approaches and the preference concept to built new system able to assist stakeholders in the decision-making problem.

These overall optimization results indicate that this new IDSS has the capability in handling various complex selection problems and can offer optimum solutions under lower computational efforts.

## REFERENCES

- [1] I.A. Stanescu, A. Stefan, D. Stefan, F. Dragomir, N. Olariu and O. Dragomir, "Intelligent decision support for renewable energy providers", Proc. Control, Decision and Information Technologies (CoDIT), 2014 International Conference on, pp.488-492, 3-5 Nov. 2014, doi: 10.1109/CoDIT.2014.6996942.
- [2] D. Arnott and G. Pervan, "A critical analysis of decision support systems research revisited: the rise of design science" Journal of Information Technology, vol. 29, 2014, pp. 269–293.
- [3] D.J. Power, F. Burstein, and R. Sharda, "Reflections on the past and future of decision support systems: perspective of eleven pioneers", Decision Support, Annals of Information Systems, vol. 14, 2011, pp.25-48.
- [4] J. Figueira., S. Greco and M. Ehrgott, "Multiple Criteria Decision Analysis: State of the Art Surveys", Springer, 2004.
- [5] S. Liu and P. Zaraté, "Knowledge based decision support systems: a survey on technologies and application domains group decision and negotiation", A Process-Oriented View Volume 180 of the series Lecture Notes in Business Information Processing pp 62-72.
- [6] Y. De-Smet, and L. Montano-Guzmán, "Towards multicriteria clustering an extension of the k-means algorithm", European Journal of Operational Research, vol.158, 2004, pp. 390–398.
- [7] A.Dadda and B.Ouhbi, "A new hybrid MCDM-Method: Application to renewable energy project management", Proceeding IEEE, IESM, Morocco, 2013.
- [8] A-H.I. Lee, H-H Chen and H-Y Kang, "Multicriteria decision making on strategic selection of wind farms", Renewable Energy, 2009, vol. 34, pp. 120–126.
- [9] F.Cavallaro, "Multicriteria decision aid to assess concentrated solar thermal technologies", Renewable Energy, 2009, vol. 34, pp. 1678–1685.
- [10] P. Haurant, P. Oberti and M. Muselli, "Multicriteria selection aiding related to photovoltaic plants on farming field on Corsica Island: A real case study using the ELECTRE outranking framework", Energy Policy, vol. 39, 2011, pp. 676–688.
- [11] J-J Wang, Y-Y Jing, Ch-F Zhang, J-H Zhao, "Review on multicriteria decision analysis aid in sustainable energy decision-making", Renewable and Sustainable Energy Reviews, vol. 13, 2009, pp. 2263–2278.
- [12] P. A Beltrán, F Chaparro-González, J-P Pastor-Ferrando and A. Pla-Rubio, "An AHP (Analytic Hierarchy Process)/ANP (Analytic Network Process)-based multicriteria decision approach for the selection of solar-thermal power plant investment projects", vol. 66, 2014, pp. 222-238.
- [13] Y. De-Smet and S. Eppe, "Relational Multicriteria Clustering: The Case of Binary Outranking Matrices", Lecture Notes in Computer Science Springer, 2009, pp. 380–392.
- [14] R. Baroudi and S. Bahloul, "Towards the definition of relations between clusters in multicriteria decision aid clustering", Procedia Computer Science, vol. 17, 2013, pp. 134 – 140.
- [15] Y.M Wang and Y. Luob, "Integration of correlations with standard deviations for determining attribute weights in multiple attribute decision making", Mathematical and Computer Modelling, vol. 51, 2010, pp. 112.
- [16] R. Mari, L. Bottai, C. Busillo, F. Calastrini, B. Gozzini and G. Gualtieri, "A GIS-based interactive web decision support system for planning wind farms in Tuscany (Italy)", Renewable Energy, vol. 36 ,2011, pp.754-76.

# IMPLEMENTATION NEAR FIELD COMMUNICATION (NFC) IN CHECKPOINT APPLICATION ON CIRCUIT RALLY BASE ON ANDROID MOBILE

Gregorius Hendita Artha K., S.Si., MCs

Faculty of Engineering, Department of Informatics  
University of Pancasila

*Abstract* - Along with the rapid development of information technology and systems that were built to support business processes, then the required transaction data more quickly and safely. Several mechanisms are now widely used transactions with NFC include Internet Online Payment, Smart Cards, Radio Frequency Identification (RFID), Mobile Payment, and others. Where the mechanism - the mechanism is designed to simplify the user make transactions whenever and wherever the user is located.

Build a new innovation from Checkpoint Apps In Rally Car circuits with Method NFC (Near Field Communication) Android Based Mobile. Basically, this is all the user system rally car competition organizers who set up several posts in the circuit for participants to be able to monitor the checkpoint that has been passed the participants are provided in each post - checkpoint.

With the demand for speed in transactions, security, and ease of getting information, so the research is to discuss the checkpoint information on the rally car circuit method NFC (Near Field Communication) based mobile android. By using NFC technology in mobile devices connected to the checkpoint transaction process will be done faster, saving, and efficient.

Application Circuit Rally Checkpoint On the Method of NFC (Near Field Communication) Android Based Mobile can monitor the riders who are competing at a distance, so the crew team from each participating teams and the competition committee can see and track the whereabouts of the car which had reached a certain checkpoint. This application can be run through the android mobile to tell him where the car. The workings of web monitoring graphs are also features that can learn from each checkpoint and rally car so that it can be used easily in view of a moving car on the racing circuit. Android apps only support the devices that already have NFC reader, as in the designation as a liaison with NFC card. All mobile applications and websites related to the wifi network that has been provided so that the system can store data and display it on a website monitoring.

*Keywords*-NFC, Near Field Communication, Android, Rally, Checkpoint

## I. INTRODUCTION

### 1.1. Background

Along with the rapid development of information technology and systems that were built to support business processes, then the required transaction data more quickly and safely. Rally car competition at many checkpoints that the driver was driving by in a live competition by passing the post - the post that has been provided by the committee. Checkpoints are not allowed to be told the race committee, until in the end participants will find the post - the post that has been provided, but for the trip will be on the committee and various instructions tell the participants of the competition must be considered in order to arrive at the checkpoint or the checkpoint. Post - post guarded by the committee to record the time of each passing car. Registrars are still manually and takes a long time. Under these circumstances turns out to have less efficiency, speed, and savings, with the support speeds up timing and accuracy of the information contained determine the success or victory of the participants and the race. With the demands for speed and ease of transaction information, so the research is to discuss the checkpoint information on the rally car circuit method NFC (Near Field Communication) based mobile android.

### 1.2. Purpose

The purpose of this paper is to build a new innovation from Checkpoint Apps In Rally Car circuits with Method NFC (Near Field Communication) Android Based Mobile. Basically, this is all the user system rally car competition organizers who set up several checkpoints in the circuit for participants to be able to monitor the checkpoint that has been passed the participants are provided in each post - checkpoint.

### 1.3. Problem

The problems of this paper :  
How to create applications using the android mobile checkpoints as checkpoint reader?  
How to monitor motorists who are competing from a distance?  
How do I work to find out the data on each checkpoint and rally car?  
How do I create a user friendly application with a checkpoint or easy to understand?  
How does NFC so they can relate to android mobile?  
How do I network to connect all of the data to be connected?

### 1.4 . Limitations of Application

Applications only for reading smart tags are already in the set . Admin Application only monitoring checkpoint information on each participant's race already past the post - checkpoint .  
Application client only collects information on the device that brought checkpoint . Not discuss security applications to share key NFC .

### 1,5 . Methodology

The methodology used in completing this application with AGILE METHOD are as follows :

### Data Collection Phase

At this stage, information retrieval and information obtained from the literature study books and material - other related materials obtained from the Internet . Collection of data obtained from various reference books that provide a source of information about NFC technology and information rules of competition rally cars .

### Phase Data Processing

At this stage, requirements analysis and system design creation and development of the application made to the programming language Java , JavaFX , MySQL Database Query , and php in favor and 3 pieces of manufacturing equipment, NFC cards , some Android Device , Wi-Fi Hot - Spot and Laptops as a Server .

### Design Application

At this stage, design process, design database and design GUI for Application on mobile phone and PC

### Implementation

At this stage, make for programming to develop application on mobile and PC.

### Test Application

At this stage, testing for application on mobile phone and PC.

## II . LITERATURE

### 2.1 . Understanding NFC

Currently , Near Field Communication ( NFC ) is one of the enablers for computer . NFC is a wireless communication technology and short-range two-way using the 13.56 MHz signal with a bandwidth of no more than 424 Kbps . NFC technology requires touching two NFC - compatible devices together for a few centimeters . [ 1 ]

### 2.2. History NFC Technology

The main vision of the integration of NFC is personal and private information such as credit card or cash card data into the phone. Therefore, security is the most important concern, and even short-range wireless communication range provided by RFID technology is considered too long. Shield is necessary to prevent unauthorized people from eavesdropping on a private conversation because even without a power source, passive tags can still be read more than 10 meters. This is the entry point where NFC technology integrates RFID and NFC contactless smart technology in mobile phones. That revolution NFC technology is illustrated in Figure below. Gray area in Figure 2.1 shows the development environment that supports NFC technology directly. [1]

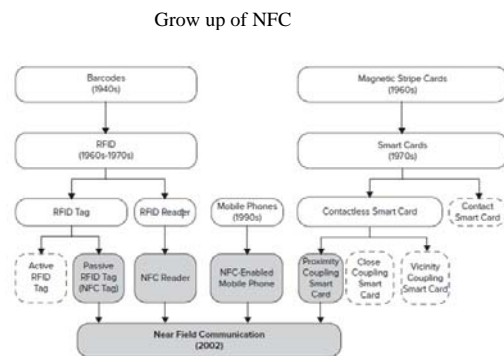


Figure 1. Grow up of NFC

### 2.3. NFC On Device

User awareness is important to do NFC communications. The first user to interact with smart objects such as NFC tags, NFC reader, or other NFC-enabled mobile phone using the mobile phone. [1]

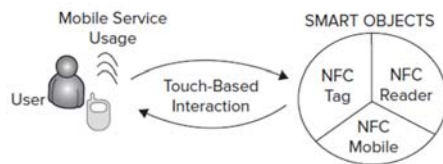


Figure 2. NFC on Device

#### 2.4. Definition of Rally Competition

At the beginning of the competition, participants get at least two sets of instructions. General instructions are the basic rules for the event, which contains definitions, priorities, and other information. Route instructions direct you throughout the course of the initial checkpoint location to the location of the finish checkpoint. Participants will not be lost if the participants just follow the route instructions in a clear manner. The other type of instruction can interact in various ways with the General Instructions, Instruction Route, and one another. [10]

#### 2.5. Checkpoint

Checkpoint is a priority that must be achieved in every rider rally participants. Each participant was required to arrive at a particular checkpoint in time that has been set. [10]

#### 2.6. Android

Android is a Linux-based operating system designed primarily for touch screen devices (touchscreen) mobile such as smartphones and tablet computers. [2]

### III. SYSTEM DESIGN

#### 3.1. Architecture Application

All android devices should connect its network to a hotspot that has been provided by the committee in order to access some features that are in need during the competition. Then throw parameters for the data from the database server to the client computer or android client can through the http request method.

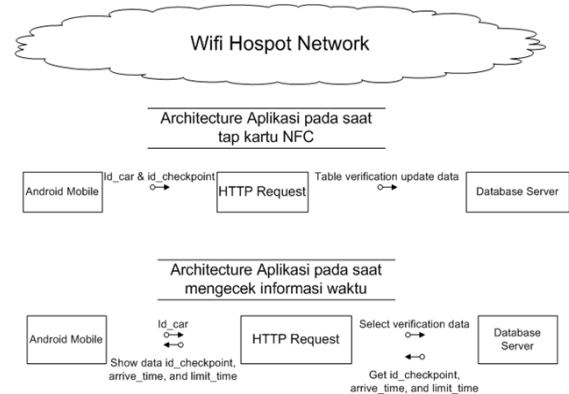


Figure 3. Architecture of Application

#### 3.2. Flow Process

Sequence of activities that can be seen in the application process checkpoint system includes image as below. All the process will be recorded on the database storage on the server, therefore monitoring for ongoing competition could be done.

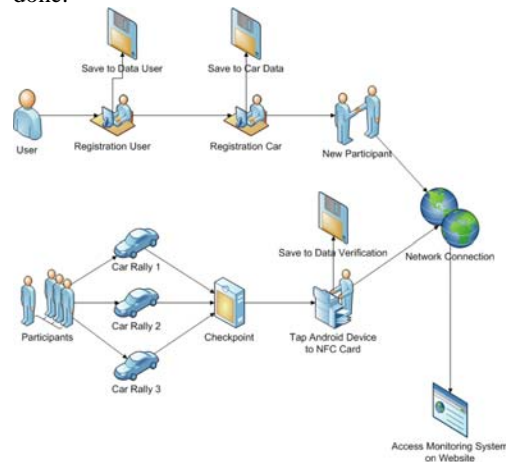


Figure 4. Flow Process of Application

### IV. IMPLEMENTATION

Implementation stage is the stage of implementation of the system that can be operated. At this stage clear about, software implementation, Hardware Implementation, Database Implementation, Implementation Program Installation, Use and Implementation Program Interface.

#### 4.1. Hardware Specifications

Table 1  
Hardware spesification on PC

No	Nama Komponen	Spesifikasi
1	Processor	Core i5 ~ 2.60GHz
2	Memory	4 GB ~ lebih tinggi
3	VGA Card	Radeon Graphic
4	Name	Specification
5	Harddisk	320 GB
6	Keyboard	Standart
7	Mouse	Standart
8	Printer	Standart
9	Monitor	15 inch

Table 2  
Hardware specification on Android Mobile

No	Name	Specification
1	Processor	1 GHz ~ lebih tinggi
2	Memory	512 MB ~ lebih tinggi
3	OS	Android OS, v4.0.3 (Ice Cream Sandwich)
4	NFC Feature	Yes
5	SD Card	No

Table 3  
Hardware specification on NFC card

No	Name	Specification
1	NFC Card	Yes
2	Type Card	MIFARE DESFire EV 1 dan MIFARE Ultralight

#### 4.2. Software Specification

Table 4  
Software specification on PC

No	Name	Specification
1	Sistem Operasi	Microsoft Windows 7
2	Aplikasi Pemrograman	Java, Java FX, Android, PHP, HTML, XML
3	Aplikasi Basis Data	My SQL
4	Aplikasi Design Mock Up	Pencils

Table 5  
Software specification on android mobile

No	Name	Specification
1	Sistem Operasi	Microsoft Windows 7
2	Aplikasi Pemrograman	Java, Java FX, Android, PHP, HTML, XML
3	Aplikasi Basis Data	My SQL

#### 4.3. Database Implementation

Creation of databases that have been arranged in the system there are several tables that relate to each other with another table. Can be seen from the picture below, which defines the relationship between the tables.

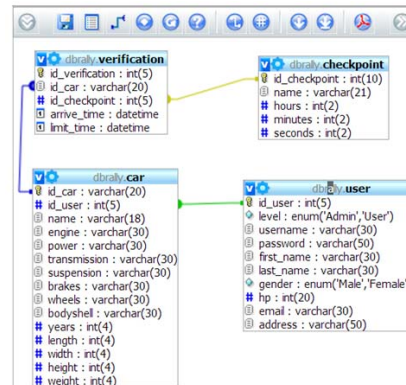


Figure 5. Table on Database

#### 4.4. Implementation

Making the menu on the website that has been created there are several options, namely: Home,



Car, User Checkpoint, Verification, Logout.

Gambar 4.6. Menu



Figure 6. Menu on Website

#### 4.5. Input Implementation

The login page is one scenario of implementation contained in the form of data input username and password. Then there are two buttons "Login" and "Reset". The title that appears as "AUTHENTICATION RALLY CAR COMPETITION SYSTEM" as the title of the login page.



Figure 7. Form Login

#### 4.6. Process Implementation

On the implementation scenario, the user is mentap android apps on NFC cards that are in a spot checkpoint. Here the mechanism of the implementation process.

On the home page there is a blank chart and a description of the car and the name of the checkpoint on the order of the list is already registered. In the detail monitoring list above shows a car which had passed the checkpoint.



Figure 8. Chart of Application

If the rider has reached the checkpoint and made it to the checkpoint, it will display the text "You're number 1 arrived in the checkpoint." On display android mobile.

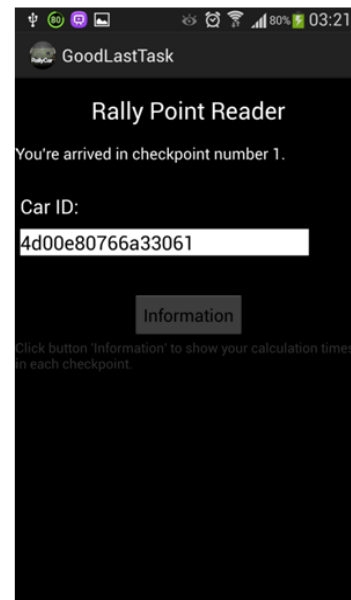


Figure 9. Car ID on Mobile Android

The last process in the home, the application will change the chart on Car ID: 4d00e80766a33061 be increased until the checkpoint number 1. Then the details of the Ford Rally 1.6 major ranks in Kemang checkpoint. Final display can be seen from the picture below.





Figure 10. Chart of Check point

#### 4.7. Output Implementation

Implementation of real output to display program information system built rally car on application program check point on the rally circuit method NFC (Near Field Communication) Android Based Mobile. For more details, the implementation of the output can be seen below.

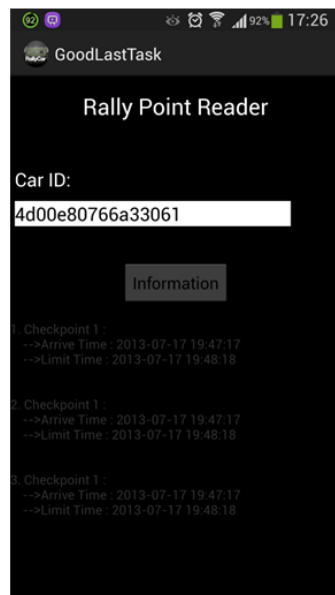


Figure 11. Rally Point Reader on Mobile Android

#### Output Car Information Data on Website

Figure 12. Output Car Information

### V. CONCLUSIONS AND SUGGESTION

#### 5.1 . Conclusion

Based on the results and discussion on the implementation can be concluded that the Application Circuit Rally Checkpoint On the Method of NFC ( Near Field Communication ) Android Based Mobile can monitor the riders who are competing at a distance, so that the competition committee and the participants can see and track the whereabouts of the car that has reached the certain checkpoint . This application can be run through the android mobile to tell him where the car . The workings of web monitoring graphs are also features that can learn from each checkpoint and rally car so that it can be used easily in view of a moving car on the racing circuit .

#### 5.1 . Suggestion

After review and evaluation so that when the perceived need to do application development . In terms of performance , innovation , ideas , not limited to the processing of data . Application development for Circuit Rally Checkpoint On the Method of NFC ( Near Field Communication ) can be a way hosting server that can be accessed through the site , so that viewers at home can also watch the passage of the ongoing competition . While on the graphic user interface in android mobile needs to be made as attractive as possible for future development .

## VI. REFERENCES

- [1] Koskun Vedat, Ok Kerem, Ozdenizci Busra, 2013, *NFC Application Development for Android*, United Kingdom, John Wiley & Sons Ltd.
- [2] Burton Michael, Felker Donn, 2012, *Android Application Development For Dummies*, Hoboken, New Jersey.
- [3] Safaat Nazruddin, 2012, Pemrograman Aplikasi Mobile Smartphone dan Tablet PC berbasis Android, Jakarta, Informatika.
- [4] Kurniawan Hendra, Mardiani Eri, Rahmasyah Nur, 2013, Aplikasi Inventory menggunakan Java Netbean, Xampp, dan iReport, Jakarta, Rumpitekno.
- [5] Ramadhani Anis, 2013, *Jurus Rahasia Pintar Menguasai Android Untuk Pemula*, Jakarta, Rumpitekno.
- [6] Jackson Wallace, 2012, *Android Apps for Absolute Beginners Second Edition*, US, Apress.
- [7] Friesen Jeff, 2013, *Java for Android Development Second Edition*, United states, Apress.
- [8] Kadir Abdul, 2009, *From Zero to a PRO*, Yogyakarta, Andi.
- [9] Siregar Michael Ivan, 2011, *Membongkar Source Code berbagai Aplikasi ANDROID*, Bandung, Gava Media.
- [10] <http://www.therallyeclub.org>, on 17 March 2013 14:04:30

# E-GOVERNMENT IN THE ARAB WORLD: CHALLENGES AND SUCCESSFUL STRATEGIES

Omar Saeed Al Mushayt  
MIS Department, Business College, KKU, Abha, KSA.

**Abstract-**Information Technology (IT) with its wide applications in all aspects of our life is the main feature of our era. It is considered as the telltale of development and progress of a country. That is why most countries are implementing IT in all areas through the establishment of the concept “e- government”. In this paper, we propose the importance of e-government, its contents, requirements, and then demonstrate the reality of e- government in the Arab World, discussing its challenges and successful strategies.

## **KEYWORDS:**

*Information Technology, e-government, e-government in the Arab World.*

## I. INTRODUCTION

Information Technology has the power to change the work pattern, administrations in all areas: upgrading performance, gaining the time, money and effort. It provides the possibility of involving citizens and civil society in the policy debate, through direct dialogues which help more understanding of citizen needs which leads to make optimal decision regarding population and this is why e-government was adopted by most countries in the world [1-15]. E-government concept has emerged, at the global level, in the end of 1995 when the central mail in Florida State, USA applied it on its administration [2-3]. But the official and political birth of e-government concept was born in Naples in Italy in March 2001. As a concept, e-government means: the exploitation of information and communication technology to develop and improve the management of public affairs by means of official government service delivery between both government agencies themselves and between clients by using the internet and Information Technology according to certain security guarantees to protect the beneficiary and the author of services which can be categorized into three levels: 1) Information dissemination in which the data and information are disseminated to public; such as data of tax statement. 2) Level in which the beneficiary to fill the tax declaration form. 3) the level where the recipient to pay the tax. As example, Brazil is the first country that adopted the system of tax declaration over the internet in 1997, and by the end of 1999 about 60% of the tax permits were filled using the internet [5-7]. As a tax, many other services can be done over the internet, such as, renovation of passports, airline bookings, timing of hospitals, professions and business licenses, etc. Developing and applying e-government concept can achieve significant results at all direction: economic, political, and social, on the other hand, it can be considered as the way that responds to the aspirations of beneficiaries, institutions and individuals through providing better services and can melt the ice of complexity of bureaucratic and routine procedures, and provide access to all services and meet the needs of citizens on the basis of fairness, equality and transparency. Moreover, it is the way to activate

the government machinery and develop its performance and ease the administration burden on it. It is the best way to store the confidence to the citizens in the administration. In fact, e-government could be described as revolution in thinking and implementation, revolution to eliminate the waste of time, effort and resources. However, this advanced technique requires a lot of work such as the establishment of steady, secure and advanced computer communication system, appropriate legislative framework and suitable organizational structure for the electronic government system. In addition to dissemination of digital culture with the awareness rise of importance of the concept of e-government and its benefits for citizens, institutions and government itself.

#### CONTENT AND SCOPE OF E-GOVERNMENT:

The concept of e-government is based on the following important elements such as: a) a compilation of all activities, interactive and responsive services which should be uploaded at one single official government portal, b) permanent interaction with the public with the ability to provide all necessary requirements regarding information services for the citizens, and c) Rapid effective liaison, coordination and performance among the government sectors.

For example, if we examine the American experience of e-government, we find that it considers most the government procurement and trade relations between governments and its sectors, institutions, public, businesses, and private sectors and that reflects the economic conditions and mental investment thinking. While, if we consider the European experience in e-government, we find it considers the most the consumer and works to protect the service of citizen.

The developing countries go in the middle of these two models, but if we want a real effective successful e-government, we have to consider the essential structure of e-government, the performance of governmental agents and the general culture of citizens.

To illustrate the content of e-government, one can access the portal of e-western government that announced that the compilation of building e-government such as USA ([www.usa.gov](http://www.usa.gov)). It has simple appearance with gateway to the three branches of e-government: executive, legislative and judiciary which are the available to all entrance of the institutions, departments of each authorities, and organizations. We can find easily integrated electronically services in all fields: health care, social security and personal status, immigration, taxation, business, investment, study and..etc. Moreover, there is the availability for electronic payment to the services sectors which require some fees for certain services. There are also real-time electronic forms which can be filled and submit digitally, availability of search any governmental information using giant advanced research engine, vast amount of information are available and many important links are also available. And this really a big challenge that all information are available within the technical standards and real-time access to the internet. So, e-government building means, taking into account all practice governmental activities in the actual reality must be emittated and uploaded online.

#### *E-government Building Requirements*

There is a need for many types of requirements for building e-government which can be categorized into Technical requirements, Regulatory requirements, Administrative requirements, Legal requirements, and Human requirements. To fulfill all of the listed requirements, we have to:

- Solve the problems existing in the real world before moving to the electronic environment; the government must provide all necessary information for the citizens through the internet and there must be a policy which identify all information, documents and governmental models directly through the internet and any new governmental document must be uploaded on the internet. In this context, the most important problem faces us is the documentation, since there is effective system for documentation which can upload governmental documents in its place and in real-time. It is very dangerous to build e-government before solving such important problems.

- Provide the appropriate structure and strategies for building societies which requires creation of electronic interactive mediator between the governmental institutions and citizens, so as to provide information directly from the event of any commercial operation has been performed earlier, in addition to use the video conferencing to facilitate the communication between the citizen and government employee. E-government must reflect the government's quest to reinvent itself in order to perform its functions effectively in the global economy market. E-government is the radical shift in the way of accomplishing its deals since the beginning of industry era.
- Solve all problems related to the legality of commercial exchanges with providing its technical and organizational means, since all exchanges dealing with money must be uploaded on the internet. Such as, the possibility of payment bills, various government fees and many other which really need to be discussed according to the law and the acceptance of electronic payment instead of cash payments. In addition to protect the security of electronic communication, respect the privacy and formulate the law of criminal accountability of the internet thieves i.e. e-government requires a new effective and appropriate which can work perfectly in the field of governmental administration.
- We must draw a successful plan for e-government; identify the applicable goals with management which can stimulate the investment opportunities and to treat all levels in a realistic and transparent way with feedback and analysis of what we have achieved and need to be fulfilled in order to insure the availability of the required elements of development.

*The Benefit of E-Government:*

- The application and implementation of e-government have many benefits including:
- The efficiency of procedures and rationalization of cost:
- With the evolution of the delivery of government services and streamlining regulations and procedures, the government can reap many benefits of e-government for its institutions and private sectors, such as:
- Raise the level of performance: increase the speed and accuracy of information transmission between the various governmental departments which reduces the duplication of data entry and facilitates the access to information from the commercial sectors and citizens. The procedures of recycling electronic information could be completed within standard time.
- Increase the accuracy of data: Due to availability of access to the requested information from the initial point of entry. The mutual confidence in validity of data is high and the errors of manual entry are reduced.
- Reduced the administrative procedures: the availability of information with the digital format, reduces the paperwork and mobilization of data manually and the availability of electronic submission.
- The optimal use of human energies: the digital information can be easily moved, manipulated, disseminated and reused.
- Recognized public services: e-government is dedicated to improve the quality of public services, streamline the governmental procedures and ease the commercial and business dealings and transactions. It increases the capacity of networking and connectivity with emphasis on the access of services to all regions and sectors effectively.
- The growth of business: e-government is more than just e-commerce; e-trade deals with the sale of goods or services through the use of technology, while as the e-government is concerned first and foremost the use of technology to raise the level of governmental services, coordinate the various governmental departments

and sectors to ensure the benefits of citizens and companies and the government itself. These services include: government procurement, bidding for goods and services, registration and renewal of licences and permits, creation of jobs and payment of dues, etc., thus the e-government aims to transform the deal between its sectors, business sectors, and citizens. It is the greatest supporter to economy and it plays a key role in recruitment of human cadres and it can be the catalyst for growth of the information Technology in the state and can urge the adoption of IT in all sectors of economy which can support efforts to attract foreign investment and upgrading the capabilities of business sectors to compete globally.

#### E-GOVERNMENT IN THE ARAB WORLD:

We demonstrate the reality of e-government in the Arab World and highlight the most important practical steps to promote the Arab presence in the information society and establishing the foundation of e-government as our time is the age of information and communication technology which has impacts in all areas, cultural, economy, military and social development. So, many countries increased the interest of implementing e-government. Some countries increase the spending on the services centers to improve e-government, such as USA which spent 6.2 billion dollars in 2003, and the UK spent 4 billion dollars for building e-government in various institutions and this model will be used for European countries. This amount of spending has been supported by political government sectors in many countries in the world to overcome the problem of bureaucracy and centralization and to save the time and efforts.

USA is ranked first followed by Australia and New Zealand, Singapore, Norway, Canada and United Kingdom, where the index was adopted many elements of the quantity that can be measured such as the ability of population in all parts of the state to access information electronically. The index reflects the overall economic arrangement of the states, and therefore the outcome of the report has emerged the significant relationship between the economic development of the state and the effectiveness of e-government; there is a lack of coordination between governmental organizations with regard to building e-government and there is a digital divide between the institutions responsible for the public administration.

We have considered a number of studies on e-government in the Arab world and observed the followings:

- There is a digital divide between the Arab governments regarding to the application of information infrastructure; it is clear from the content of number of e-government portals.
- Lack of awareness of all technology elements; hardware and software, which is very important for building e-government; many of Arab portals are tuff without contents which can serve citizens.
- There is no access to published literature and intellectual productions in the area of technology.
- There is no clear relationship between Arab Portals of each Arab state and the application of electronic government projects and the ability of sites to provide the needed services to citizens.
- There is a need for all Arab States for more efforts to construction government positions, both in terms of form or substance.
- There is a strong relationship between the simplification of procedures and laws of the state and the states' ability to build e-government projects.
- The political rhetoric of political leaders has an effective influence on the building of e-government and can enhance the relations with the Arab citizens.

Thus, to devote the Arab presence in the information society and electronic government, we have to:

- Strengthen the infrastructure of information and communication technology with taking into account the geographic distribution of Arab countries to ensure the access of services to beneficiaries.
- Strengthen Arab administrations and Institutions to improve the delivery of governmental services, and reconstruction of the organizational structures of these institutions to ensure the specialized departments in information technology and enhance the governmental plans in this regard.

- Bridging the digital divide between the Arab governmental institutions inside the state so that it can provide services to Arab citizens.
- Simplify governmental procedures and reduce its number with the abolition of bureaucracy by adoption of the principle of transparency, as well as the reduction of laws and legislation which restrict the citizen.
- Promote studies in the field of e-government to shed the light on the criteria of measuring electronic governments enterprises and to evaluate the accomplishment of e-government projects in the Arab world.

#### CONCLUSIONS AND RECOMMENDATIONS:

- We can deduce from the above, that e-government in the current pattern has not yet reached the desired system which needs great development in many aspects, since it is not only shifting from a simple actual system to electronic ones, but it is a full complicated automated and interrelated system. In addition, the development of such system could result in some negative aspects that must be dealt with great caution. It is the major challenge of real government which can face all information and cultural invasions existing around the world. So as to build such e-government, we recommend the followings:
- There is a need for understanding of different types of e-government components and their requirements; to activate the positives and reduce the negatives.
- Non-importation of ready-made templates for e-government: we must construct and apply an appropriate system for our Arab societies due to the differences in the infrastructure, circumstances and factors that constitute each component of e-government in the Arab World.
- Eliminate the problems of computer illiteracy and spread the digital knowledge in the Arab World before the application of e-government.
- Study and evaluate all the negatives that arise in the process of applying e-government, such as the problems of unemployment and privacy and attempt to find optimal solutions for them in advance.
- Activate the role of private sectors in the process of transition to e-government to ease the burden of government, as well as the provision of skilled labor in the field of Informatics and upgrading the capacity of public to deal with these new technologies.
- Formulation of computer and communication workshops in all departments and government sectors to analyze, develop and unify the existing infrastructure with consolidation of database and software applications.
- Appropriate financial support to cover all the technical and software costs in all sectors.
- Finally, with the consolidation of efforts, dedication to work and coordination, we can achieve the desired goals and catch up with others of the first journey.

#### REFERENCES

- [1] Layne, K., & Lee, J. (2001). Developing fully functional e-government: A four-stage model. *Government Information Quarterly*, 18(2), 122–136.
- [2] Sharma, S. K., & Gupta, J. N. D. (2002). Transforming to e-government: A framework. *Proceedings of the 2nd European conference on e-government*, U.K. Oxford, 383–390.
- [3] Bollettino, J. (2001). E-government-digital promise or threat? *Oil and Gas Investor*, 5.
- [4] Ronaghan, S. (2002). Benchmarking e-government: Assessing the UN member states. *United Nations Division for Public Economics and Public Administration and American Society for Public Administration*.
- [5] Jupp, V. S. S. (2001). Government portals—The next generation online. *Proceedings of the European conference on e-government*, 217–223.
- [6] Lau, E. (2001). Online government: A surfer's guide organization for economic cooperation and development. *The OECD Observer*, (224), 46–47.

- [7] Wauters, P. K., Hugo (2002). Web based survey on electronic public services. Cap. Gemini Ernest and Young.
- [8] Tedesco, N. (2001). Buying into e-government: Who, what and how. Summit, 4, 10–12.
- [9] ITERI. (2002). Information technologies usage research-2000. Information Technologies and Electronic Researches Institute
- [10] Bannister, F., & Neasa, W. (2002). E-democracy: Small is beautiful. Proceedings of the European Conference on e-government, pp. 49–65.
- [11] Mariam AlNuaimi, Khaled Shaalan, Moza Alnuaimi, Klaitheem Alnuaimi (2011), Barriers to Electronic Government Citizens' Adoption. 2011 Developments in E-systems Engineering.
- [12] <http://portal.www.gov.qa/wps/portal>. Last seen 30/08/2015.
- [13] <http://www.saudi.gov.sa/wps/portal/>. Last seen 30/08/2015.
- [14] <http://www.egypt.gov.eg/>. Last seen 30/08/2015.
- [15] <http://government.ae/en>. Last seen 30/08/2015.



# C-ODMRP: A Density-Based Hybrid Cluster routing Protocol in MANETs

Yadvinder Singh

Department of Computer Science & Engineering  
Sri Sai College of Engineering & Technology  
Amritsar, India

Kulwinder Singh

Assistant Professor  
Department of Computer Science & Engineering  
Sri Sai College of Engineering & Technology  
Amritsar, India

**Abstract**—Cluster based on demand multicasting provides an efficient way to maintain hierarchical addresses in MANETs. To overcome the issue of looping in the ad hoc network, several approaches were developed to make efficient routing. The challenge encountered by multicast routing protocols in this ambience is to envisage creating a cluster based routing within the constraints of finding the shortest path from the source and to convert a mesh based protocol into Hybrid. This paper represents a novel multicast routing protocol C-ODMRP (Cluster based on demand routing protocol), a density-based hybrid, which is a combination of tree-based and mesh-based multicasting scheme. K-means algorithm approach also used to choose the *Cluster Head*, which helps in dynamically build routes and reduces the overhead of looping. C-ODMRP is well suited for ad hoc networks, as it choose *Cluster Head* through shortest path and topology changes frequently.

**Keywords**—C-ODMRP, *Cluster Head*, K-means, density-based Hybrid, *Route Table*, MANETs.

## I. INTRODUCTION

A Mobile Ad-hoc Network (MANET) is a temporary wireless network composed of mobile nodes that dynamically self-configures to form a network without any fixed infrastructure or centralized administration [17]. On Demand Multicast Routing Protocol (ODMRP) is a multicast routing protocol for mobile ad hoc networks. It uses the concept of "forwarding group," [17][26] a set of nodes responsible for forwarding multicast data, to build a forwarding mesh for each multicast group. Its efficiency, simplicity, and robustness to mobility furnish by maintaining and using a mesh instead of a tree. Several routing schemes have been proposed for the purpose of providing adequate performance and node movement patterns. The reduction of channel overhead and the usage of stable routes make ODMRP more scalable for large networks and provide robustness to host mobility. There were some drawbacks in ODMRP such as Short term disruptions such as jamming, fading, obstacles and Medium term disruptions, e.g. FG node moving out of field [17].

Mobile ad hoc network routing is classified as proactive [26] in which each node in the network has routing table which contain the information of broadcasting of data packets. At present time, to retain stability each station broadcast and modify its *Routing Table* time to time. Reactive routing protocol lowers the overhead as it routes on demand. It uses the concept of

flooding (global search) the *Route Request* (RREQ) for route discovery on demand by sending the packets throughout the network.

The network will undergo too much routing overhead wasting valuable resources, if it is too high. Thus ODMRP cannot keep up with network dynamics, If it is too low [2].

The primary goal of an ad hoc network routing protocol is provide correct and efficient route establishment between a pair of nodes so that messages may be delivered in a timely manner. Route construction should be done with a minimum of overhead and bandwidth consumption [7].

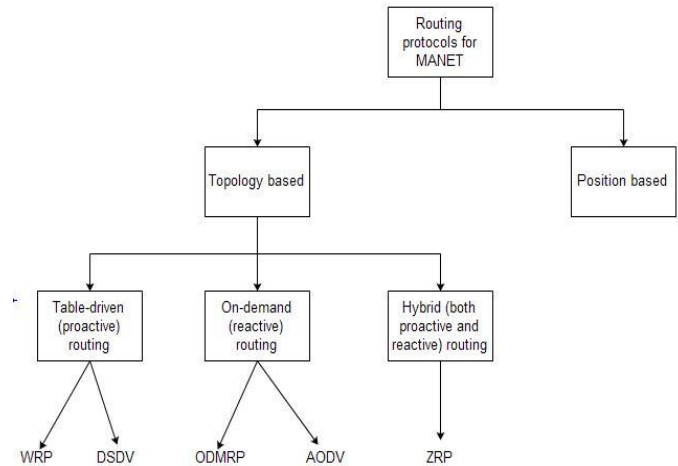


Figure 1: Classification of Ad hoc Routing protocols [3]

Multicast tree structures are fragile and must be readjusted continuously as connectivity changes. Furthermore, typical multicast trees usually require a global routing substructure such as link state or distance vector. The frequent exchange of routing vectors or link state tables, triggered by continuous topology changes, yields excessive channel and processing overhead [26].

Tree-based schemes establish a single path between any two nodes in the multicast group that are also bandwidth efficient.

However, as mobility increases, the entire tree will have to be reconfigured. When there are many sources, one may have to maintain multiple trees resulting in storage and control overhead. As a conclusion in a high mobile scenario, mesh based

protocols exceeded tree-based protocols. Examples of tree-based schemes include ad hoc multicast routing protocol (AMRoute), ad hoc multicast routing utilizing increasing ID-numbers protocol (AMRIS), and multicast ad hoc on-demand distance vector routing protocol (MAODV) [6][26][8].

Mesh-based schemes establish a mesh of paths that connect the sources and destinations and packets are distributed in a mesh structures. They are more efficient to link failures and have the high robustness as compared to tree based protocols. The major drawback is that mesh-based schemes provide redundant paths from source to destinations while forwarding data packets, resulting in reduced packet delivery and increase in control overhead. Some examples of mesh-based schemes are (a) on demand multicast routing protocol (ODMRP[26]), (b) forwarding group multicast protocol (FGMP), (c) core assisted mesh protocol (CAMP), ((d) neighbor supporting ad hoc multicast routing protocol (NSMP), (e) location-based multicast protocol, and (f) dynamic core-based multicast protocol (DCMP) [7][22].

We propose a Density based Hybrid Cluster routing protocol which combines the properties of tree-based scheme and mesh-based routing scheme. In this density based Hybrid clustering approach connection to the nodes will be tree-based and Packet Relaying will be Mesh-based. For clustering the K-means algorithm is used to create cluster based on the erupted propagation of the number of forwarding nodes.

Our results shows that the proposed heuristic Density based Hybrid Cluster, when implemented into ODMRP, it becomes Cluster based on demand routing protocol.

## II. RELATED WORK

Due to the increasing importance of Cluster based routing various multicast protocols in MANETs along with the challenges and issues existing in the MANETs. Proactive and reactive approaches then use individually lead to packet delay and routing overhead problem. Elizabeth M.Royer *et.al.*, (1999) gave a review that the primary goal of such an ad hoc network routing protocol should be correct and efficient for route establishment between a pair of nodes. They provide descriptions of several routing protocols schemes proposed for ad hoc networks they also provide a classification of schemes according to the routing strategy [7]. According to Jane Y.Yu *et.al.*,(2006) the Cluster\_Head is responsible for maintaining local membership and global topology information. Thus the inter-cluster level information is maintained by Cluster\_Heads via a proactive method [24]. According to Sung-Ju Lee *et.al* (2005), ODMRP is well suited routing protocol as it is mesh-based rather tree-based and also used the concept of forwarding group to multicast packets via scoped flooding. ODMRP is effective and efficient in dynamic environments and scales well to a large number of multicast members [26]. Neha Gupta *et.al.*,(2012) conclude that the primary goal of an adhoc network routing protocol is to provide efficient route between a pair of nodes so that messages may be delivered in a timely manner & the route construction should be done with a minimum of looping overhead and bandwidth consumption. They focused on cluster-based routing on demand protocol and uses

the clustering structure [3]. S. Rajarajeswari *et.al.*,(2015) performed a survey that classifies the multicast routing protocols routing structures: tree-based and mesh-based. Their study showed that multicast routing protocol may improve network performance in terms of delay, throughput, reliability or lifetime [26]. The k-means algorithm scheme can improve the computational speed of the direct k-means. The main confront lies in applying multicast communication to the scenario in which mobility is unlimited and also where frequently failures occur.

## III. CLUSTER FORMATION AND CLUSTER MAINTENANCE

### Cluster Formation

To the best of our knowledge Clustering is a well-known technique for grouping nodes that are close to one another in a network. Most of the cluster-based routing algorithms tend to use proactive approaches within the cluster and reactive routing for inter-cluster routing [24]. However, when majority of nodes present outside the cluster this type of scheme may incur significant route delay and looping overhead. The concept of clustering is to divide the *k-size* large network into *n* number of sub-networks. Any node can become a Cluster\_Head if it has the essential functionality, such as processing and transmission power. Cluster\_Head finds the Node registers with the nearest shortest distance and becomes a member of that cluster.

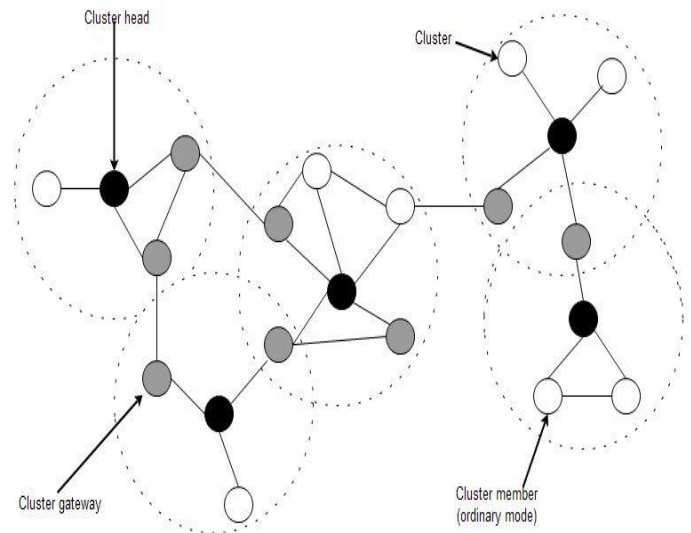


Figure 2: Cluster structure illustration

Adopting the clustering approach with ODMRP can make fewer connections existing between different zones in the network cluster such as intra-cluster link to connect nodes in a cluster and inter cluster link to connect clusters [14] [3].

### Cluster Maintenance

A member replies back a message to its Cluster\_Head when a Cluster\_Head periodically broadcast a message in order to maintain membership of a cluster. Certain conditions need to be followed such as a member might not get message from its original Cluster\_Head but from other Cluster\_Heads in that case it will join a new cluster with the shortest distance to the

new Cluster\_Head. Further, The member entry will be updated and the Original Cluster\_Head will delete it [6][8][14].

Our goal is to design a routing protocol that benefits both route delay and looping overhead. That means to use both the proactive approach as tree based scheme and reactive approach as the mesh based scheme to make route delay bearable and the number of control packets should be controllable as well.

### K-means Density based Hybrid Clustering

When the network topology changes the use of independent dominating sets as Cluster-Heads is problematic. In particular, one of the Cluster\_Head must defer to other in order to trigger Cluster\_Head changes that may disseminate throughout the network, such an effect is called chain reaction [21]. This chain reaction effect does not occur while relaxing the independence condition on dominating sets.

In this paper we presented Density-Based Hybrid clustering with the help of K-means algorithm which works as follows:

- i. K-means Clustering algorithm will group large network into  $n$  number of small sub networks.
- ii. A centroid will be generated in each and every sub-network.
- iii. Distance from the centroid to all the nodes will be calculated.
- iv. The minimum distance node will be selected as Cluster\_Head.

### IV. C-ODMRP ROUTE DISCOVERY AND HYBRID CREATION OPERATION

After finding the centroid using density based K-mean clustering the minimum distance node will be selected as the Cluster\_Head and the C-ODMRP route and Hybrid creation operation starts.

1.  $S$  floods a *Join Query* to entire network to refresh membership.
2. During *Join Query* it will create mesh between source and all **Cluster\_Heads**
3. After that Query will reach to multicast destination
4. In *Join Reply* phase multicast destination sends *Join Query* back to source through shortest path.
5. Data will be forwarded to the same path from where the *Join Query* came.
6. Data will be forwarding from source to **Cluster\_Head** in which multicast Destination is there.
7. *Reply Phase* broadcast **Cluster\_Head** to multicast destination and an acknowledgment will be send back to **Cluster\_Head**
8. *Join Reply* is propagated by each Forwarding Group member until it reaches source via shortest path.

9. Routes from source to all **Cluster\_Head** build a *Tree*, then all cluster head joins with each other through a *Mesh-Based* scheme which gives the composite solution as a *Hybrid cluster*.

### Multicast Route and membership Maintenance

#### Route Table

A Route Table is maintained by each and every node and created on demand. All entries are updated or inserted when a non-duplicate JOIN REQUEST is received in route table. Also Routing table provide us the information about transmitting Nodes and store the information about which node or hop act as source, destination, intermediate hops in routing routine[4][26][9].

#### Forwarding Group

It is the subset of nodes which forwards multicast data packets via scoped flooding. Data is delivered by this forwarding group. Nodes that are having shortest paths will be selected as the forwarding group and will lead to make a forwarding mesh for the multicast group. This will dynamically build routes destined to the associated multicast group [26][17].

#### Data Forwarding

A multicast source can transmit packets to receivers after the group establishment and route construction process, via selected routes and forwarding groups. When it receives a multicast data packet a node forward it if it's not a duplicate one and FG\_FLAG not expired. This process minimize the overhead [17][26].

### V. DENSITY BASED HYBRID CLUSTERING PROTOCOL

We prefer to maintain the knowledge of full network topology ,but wish to avoid the inefficient flooding mechanisms. To make measurable progress in the field of MANETs routing density based hybrid routing is necessary. The methodology used in which an ad hoc network will be created and coordinates of all the nodes will be discovered. Therefore, we develop our Density based Hybrid routing protocol C-ODMRP based on the clustering scheme described in the previous section.

The idea behind the use of Hybrid cluster routing was the hierarchical structure, so single point node failures can be reduced by routing in a hybrid cluster. The availability of route always depend upon the location of the destination.in Hybrid Clustering approach the traffic volume mostly lower than proactive and reactive approaches. Periodic updates used inside each zone or between the gateways of the cluster. Usually more than one path may be available due to the hierarchical structure and the size of cluster may become large. The delay level for most of the local destination is small in hybrid approach [21][27][28].

K-means Clustering algorithm will group large network into  $n$  number of small sub networks. A centroid will be generated in each and every sub-network [6][10]. Distance from the centroid to all the nodes will be calculated. The minimum distance node will be selected as Cluster\_Head.

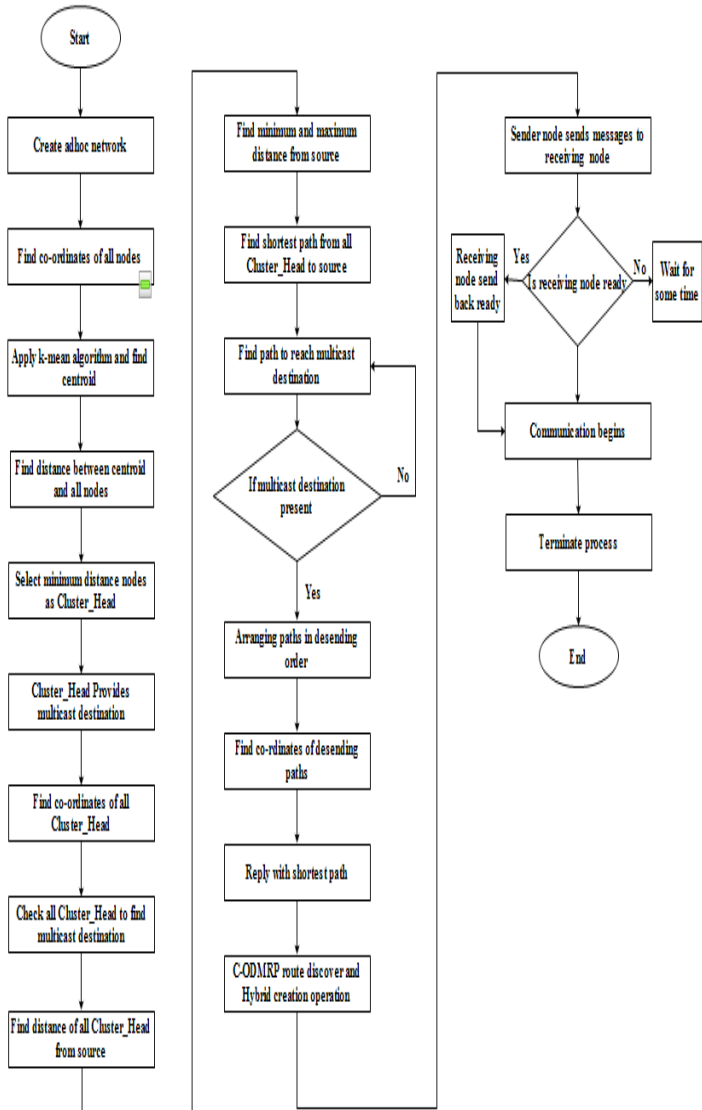


Figure 3: Density based hybrid cluster routing scenario in MANETs

After determining each node's final value, nodes send their values to their Cluster\_Head. Consequently, all nodes know their Cluster\_Head and their own values. Source node of a domain set send messages to the Cluster\_Head and further that message broadcasted to the members of that particular cluster. The Cluster\_Head will provide multicast destinations and distance of all Cluster\_Head from the source will be discovered. Minimum and maximum distance will be calculated from the source node to all the nodes present in cluster. In the further process shortest path from all the Cluster\_Head to source as shown in the Figure 4.

As there is no direct shortest path from Node 6 to Node 1 so the density based K-means algorithm will help to find the Cluster\_Head with the minimum distance. send data from source to cluster head via shortest path.

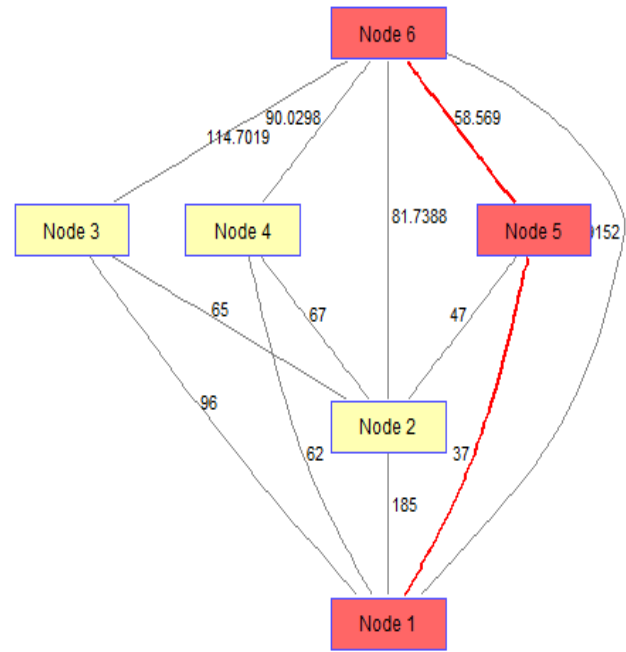


Figure 4: Intermediate based shortest path from the Source to node 1.

Sending data from cluster head to multicast destination. A reply will come from destination to cluster head and it creates mesh and send queries. Group membership and multicast routes are established and updated by the source on demand. While a multicast source has packets to send, it periodically broadcasts to the entire network a member advertising packet, called a JOIN REQUEST [26].

Similarly minimum distance paths discovered from source to all the nodes as shown in Figure 5:

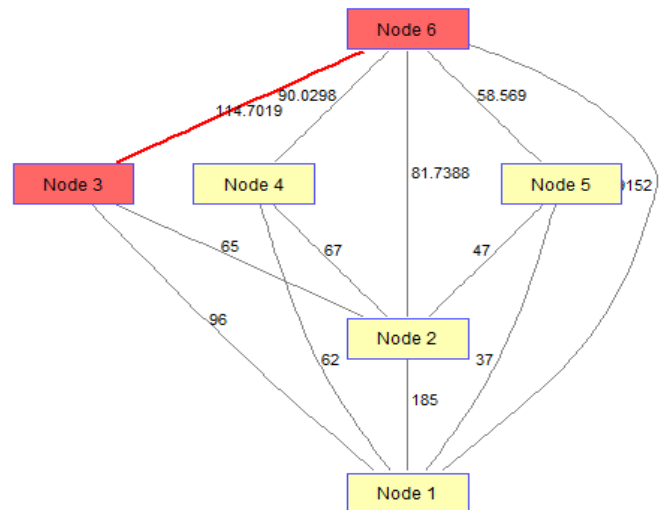


Figure 5: Direct path shortest distance from source to Node 3.



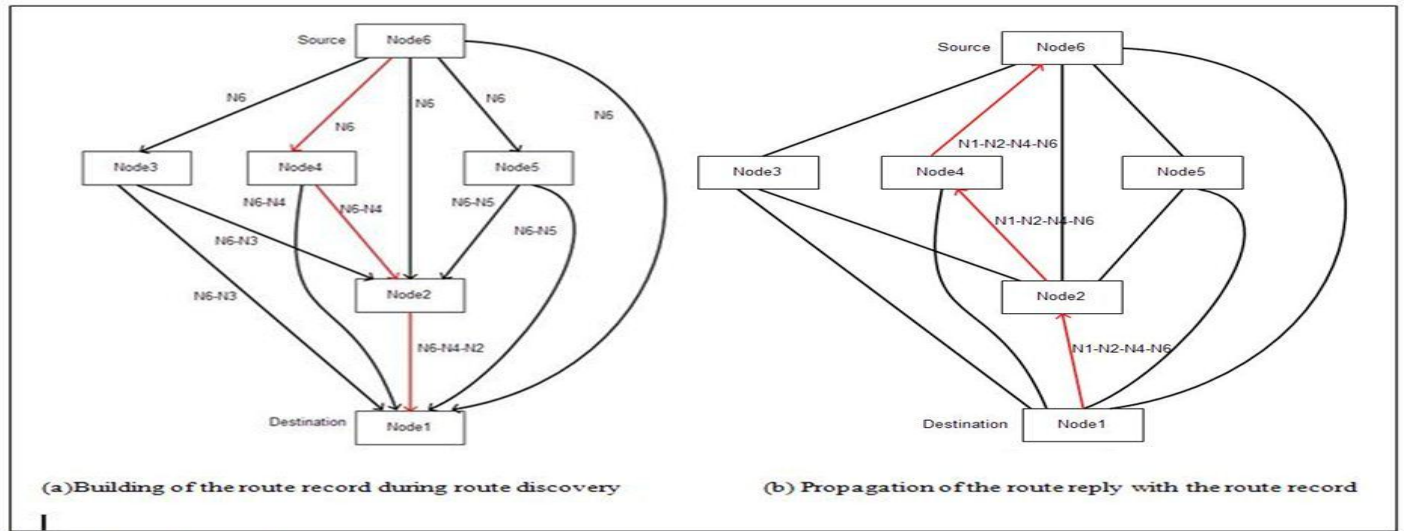


Figure 6: Creation of the Route Record in C-ODMRP

Source node starts the route discovery on demand basis of information stored as source address, destination address, and intermediate nodes addresses along the packet. It checks the route table to identify active route to the destination when a source node have some data packets to send to the destination. If the route is not present it starts the route discovery initiate. In HCR, the route discovery as of inter-cluster route discovery and intra-cluster route discovery. In Inter-cluster route discovery node send a *cluster list*

*request* (CLREQ) to its host cluster heads. After a CH receives a CLREQ, it sends back a *cluster list reply* (CLREP) to the CLREQ initiator node. after get replay source node checks is message valid If yes , the node will update the route information in its route table. Else retry to send another CLREQ for MAX\_CLREQ times. In Intra-cluster route discovery a node send packet to a destination node that locates within the same cluster. A node receives a *route request* (RREQ), it checks whether it reply to the RREQ. The node act as the requested node and an intermediate node has an active route in its routing table. Send a *route reply* (RREP) to the RREQ initiator and route Information else the RREQ is re-broadcast by the node [7][24].

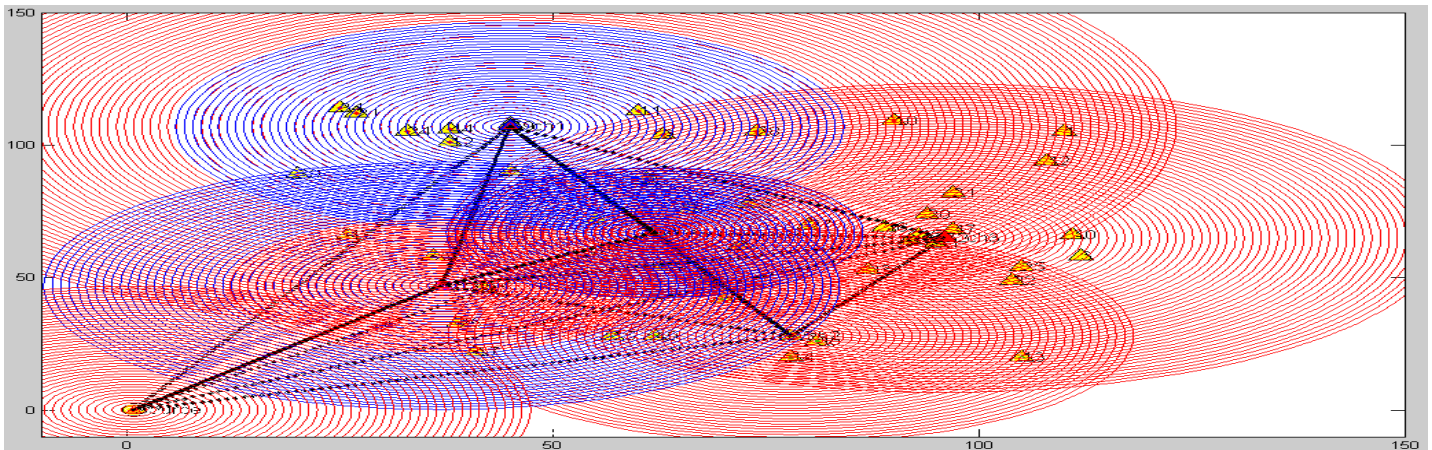


Figure 7: The overall process of C-ODMRP route discovery and Hybrid based cluster routing operations

When CLREQ initiator receives a CLREP, the node will fill the cluster list into the corresponding field in data packet's header. In this way, the packet is routed to the destination by RREP. When a node receives the packet, it will forward send the packet to the next cluster along the *cluster list*. Thus, the packet is forwarded cluster by cluster protocol, every time a source floods a JOIN REQUEST. The process continues until reaching the multicast receiver node. Once the receiver node received the JOIN REQUEST. It will declare it's joining by broadcasting JOIN REPLY message to the multicasting group [10]. If the multicast destination is present then *Route table* will arrange all the paths in descending order and start discovering the coordinates of

until it arrives at the last cluster as the destination node. After that the packet will be forwarded to the destination node by node within the last cluster. To start sending multicast data packets using C-ODMRP, if there is nodes wants to join to the multicast group, it uses JOIN QUERY. Using of JOIN REPLY will be activated when the receiver node accept to receive the multicast data packet. In C-ODMRP nodes. So in this manner C-ODMRP density base hybrid clustering algorithm works. Its uniqueness stems from the use of each multicast entry.

Due to dynamic topology of on demand multicast ad hoc networks routing is one of the challenging issues. In the past, there are various types of routing protocols used which were suitable for different situations. The density-based hybrid clustering protocol C-ODMRP have combined the advantages of both Reactive and Proactive protocols.

## CONCLUSION

This paper represented a novel multicast routing protocol C-ODMRP (Cluster based on demand routing protocol), a density-based hybrid, which is a combination of tree-based and mesh-based multicasting scheme. K-means algorithm approach also used to choose the *Cluster\_Head*, which helps in dynamically build routes and reduces the overhead of looping. C-ODMRP is well suited for ad hoc networks, as it choose *Cluster\_Head* through shortest path and topology changes frequently. The latency is decreased by using proactive protocol and the looping overhead is decreased by using reactive protocol outside the zone. Hence a C-ODMRP is a protocol presented which improves the performance of network by using the advantages of both reactive and proactive protocols. The approach is well suited for improving the time efficiency. In the past, due to looping to detect next hop or to jump to next node was not possible and took longer time. So this approach is feasible to improve the time efficiency and get rid of looping overhead. Various improvements of the protocols still in progress and will be reported in the upcoming paper.

## REFERENCES

- [1] Alpar Juttner, A. a. (2004, September 9). Tree Based Broadcast in Ad Hoc Networks. 1-21.
- [2] Alsukour, K. A. (n.d.). Novel Protocols for Improving the Performance of ODMRP and E-ODMRP over Mobile Ad hoc Networks.
- [3] Anuj K. Gupta, H. S. (2011). Review of Various Routing Protocols for MANETs. *International Journal of Information and Electronics Engineering* , 1 (3), 251-259.
- [4] Charles E. Perkins, E. M. (2015, September 30). Performance Comparison of Two On-Demand Routing Protocols for Ad Hoc Networks. *IEEE Personal Communications* , 16-28.
- [5] CHONG, J. Y. (2005). A SURVEY OF CLUSTERING SCHEMES FOR MOBILE AD HOC NETWORKS. *IEEE Communications Surveys & Tutorials* , 7 (1), 1-48. Curt Cramer, O. S. (2004, January ). Demand-Driven Clustering in MANETs.
- [6] Damianos Gavalas, G. P. (n.d.). Clustering of Mobile Ad Hoc Networks: An Adaptive Broadcast Period Approach. 1-7.
- [7] Elizabeth M. Royer, C.-K. T. (1999, April). A Review of Current Routing Protocols for Ad Hoc Mobile Wireless Networks. *IEEE Personal Communications* • April 1999 , 1070-9916.
- [9] Fahmy, O. Y. (n.d.). Distributed Clustering in Ad-hoc Sensor Networks: A Hybrid, Energy-Efficient Approach. 1-12.
- [10] Khaled Alsabti, S. R. (n.d.). An Efficient K-Means Clustering Algorithm. 1-6.
- [11] Meenakshi Diwakar, S. K. (2012). AN ENERGY EFFICIENT LEVEL BASED CLUSTERING ROUTING PROTOCOL FOR WIRELESS . *IJASSN* , 2 (2), 55-65.
- [12] Mohammed R. BAKER, M. A. (2011). A Survey of Multicast Routing Protocols in Ad-Hoc Network. *GU J Sci* , 24 (3), 451-462.
- [13] Nagarajan, S. G. (2015). An Improved Stable Multicast Routing Scheme for Misbehavior Detection in MANET. *Middle-East Journal of Scientific Research* , 23 (7), 1482-1491.
- [14] Neha Gupta, E. M. (2012). Cluster Based on Demand Routing Protocol for Mobile Ad Hoc Network. *IJERT* , 1 (3), 1-6.
- [15] Pang, I.-S. H.-H. (2007). Energy Efficient Clustering Technique for Multicast Routing. *IJCSNS* , 7 (8).
- [28] Pradeep Reddy, J. G. (2015). CLAMR-AN ENHANCED BIO-INSPIRED ROUTING PROTOCOL FOR WIRELESS ADHOC NETWORKS. *IJOABJ* , 6 (4), 1-15.
- [17] Preeti.G.Sajjan, S. (2015). Efficient Multicast Routing Protocol Using Limited Flooding For Efficient Route Discovery. *International Journal of Advanced Research in Computer Engineering & Technology* , 4 (6), 1-6.
- [18] R.S.Rajesh, M. K. (2009). Performance Analysis of MANET Routing Protocols in. *IJCSNS* , 9 (2), 22-29.
- [19] Rai, R. K. (2012). A Novel Review on Routing Protocols in MANETs. *Undergraduate Academic Research Journal (UARJ)* , 1 (1), 103-108.
- [20] S. Rajarajeswari, 2. 3. (2014). Survey on Tree Based, Mesh Based and Stateless Multicast Protocols in MANET. *IJIRCCE* , 2 (3), 216-222.
- [21] Saleh Ali K.Al-Omari, P. S. (2010). AN OVERVIEW OF MOBILE AD HOC NETWORKS FOR THE EXISTING PROTOCOLS AND APPLICATIONS. *J GRAPH-HOC* , 2 (1), 87-110.
- [22] Sanneboina Prasanthi, I. B. (2012). Implementing ODMRP (on Demand Multicast Routing Protocol) in Mantes Using Stable Link Approach. *International Journal of Engineering Research and Development* , 4 (11), 27-38.
- [23] Shio Kumar Singh, M. P. (2010). Routing Protocols in Wireless Sensor Networks –A Survey. *IJCSES* , 1 (2), 63-83.
- [24] Sikamani, S. M. (2015). Quick Recovery from Link Failures using Enhanced On-demand Multicast Routing Protocol. *Research*

*Journal of Applied Sciences, Engineering and Technology* , 9 (7),  
526-530.

[25] Soon Y. Oh, J.-S. P. (n.d.). E-ODMRP: Enhanced ODMRP with Motion Adaptive Refresh.

[26] Sung-Ju Lee, M. G.-C. (n.d.). On-Demand Multicast Routing Protocol.

[27] Xiaoguang Niu<sup>123</sup>, Z. T. (2006). Hybrid Cluster Routing: An Efficient Routing Protocol for Mobile Ad Hoc Networks. *IEEE ICC* , 1-6.

[28] Abolhasan, M., Wysocki, T. & Dutkiewicz, E. (2004). A review of routing protocols for mobile ad hoc networks. *Ad Hoc Networks*, 2 (1), 1-22.

#### AUTHORS PROFILE

**Yadvinder Singh** B.tech (CSE,2012) from Lovely Professional University, Phagwara Punjab & M.Tech (CSE,2015) from Sri Sai College of Engineering & Technology, Amritsar affiliated to Punjab Technical University. Interests are Mobile ad hoc Networks, Data Warehousing & Mining, Network Security & Cryptography.

**Kulwinder Singh**, Assistant Professor in the Department of Computer Science & Engineering at Sri Sai College of Engineering & Technology, Amritsar, Punjab. He have 2 years of teaching experience. His areas of interest includes Mobile ad hoc Networks, Artificial Intelligence & knowledge based expert systems, Biometrics Security Big data Analytics & Data Warehousing & Mining

# Life time Enhancement through traffic optimization in WSN using PSO

Dhanpratap Singh  
CSE, MANIT  
Bhopal, India

Dr. Jyoti Singhai  
ECE, MANIT  
Bhopal, India

**Abstract:** Technologies used for wireless sensor network are extremely concentrated over improvement in lifetime and coverage of sensor network. Many obstacles like redundant data, selection of cluster heads, proper TDMA scheduling, sleep and Wake-up timing, nodes coordination and synchronization etc are required to investigate for the efficient use of sensor network. In this paper Lifetime improvement is an objective and reduction of redundant packets in the network is the solution which is accomplished by optimization technique. Evolutionary algorithms are one of the category of optimization techniques which improve the lifetime of the sensor network through optimizing traffic, selecting cluster heads, selecting schedules etc. In the proposed work the Particle Swarm optimization Technique is used for the improvement in the lifetime of the sensor network by reducing number of sensor which transmits redundant information to the coordinator node. The optimization is based on various parameters such as Link quality, Residual energy and Traffic Load.

**Keywords:** Lifetime, optimization, PSO, Fuzzy, RE, QL, SS

## I. INTRODUCTION

Wireless sensor network is deployed for performing many tasks such as forest monitoring, glaciers monitoring, climate, geographical analysis and data gathering etc. Increasing popularity and utility of WSN is a great attention for the Industrialist and researchers. The major areas in which research is in progress are:

- Lifetime of Network
- Reliability of Network
- Security in the Network
- Performance of the network

Various applications [1] by Giuseppe Anastasi are associated with sensors like forest monitoring, weather monitoring, fire detection, geological monitoring, and securities over international borders etc. Sensor nodes are manufactured to handle data related to single application or many applications simultaneously.

In this paper, constrained related to lifetime of network is analyzed and developed an algorithm to minimizes traffic over the network. These by saving energy consumption of sensors and enhancing lifetime of sensor network. Sensors have very little energy resource and it is needed to save energy as much as possible without significant loss of information. There are many situations when energy of sensor node is drained out such as

- Idle Listening
- Redundant traffic
- Hot Zone
- Improper Sleep and Wake-up schedule etc.

Paper is emphasized over the redundant traffic which hampers the lifetime of WSN through over utilization of energy during transmission and reception of data packets. Proposed technology saves this energy from drain out by proper management of source nodes. On the base of some parameters, few source nodes are selected for the data transmission. The Parameters for selection are Residual energy, Link Quality and Traffic Load. Finding sensor nodes having better values of these parameters is operation related to optimization process. Evolution algorithms are better for optimization, among which particle swarm optimization technique is used in this work. Selected source nodes are scheduled in their TDMA slots which utilizes Coordinated Duty Cycle mechanism.

## II. RELATED WORK

Redundancy in network is reduced using manageable duty cycle and it is proposed in paper [2] by Rashmi Ranjan Rout. Author worked and estimates the upper bounds of network life time over bottleneck zone of the network which surrounded the sink node. Energy efficient bandwidth utilization techniques reduce the traffic in bottleneck zone. Network coding is another technique used by author for improvement in network reliability. The technologies like Duty cycle and Non Duty Cycle are integrated in the network coding and analyzed their performance and lifetime with respect to duty cycle. For the Encoding operation author used:-

$$Y = \sum_{i=1}^n q_i G_i, \quad q_i \in GF(2^S)$$

Where Y is output encoded packet is transmitted with n coefficients in the network.

$q = (q_1, q_2, \dots, q_n)$  are chosen sequence of coefficient known as encoding vector, from the set  $GF(2^S)$ . A set of n packets  $G_i (i=1,2,3,\dots,n)$  at nodes are linearly encoded into single output packet. Decoding operation is performed by equation:



$$Y^j = \sum_{i=1}^n q_i^j G_i, \quad q_i \in GF(2^S)$$

They got improvement of duty cycle with network coding protocol over only duty cycle protocol. It is approved by analyzing the result with varying the duty cycle of the nodes in the network. Latency is increases due to network coding on the nodes in bottleneck zone. For the proper decoding at the sink 50% duty cycle is needed. This limits the reduction of traffic in the bottleneck zone.

In the paper [3] by Chu-Fu Wang et al., a sink relocation strategy is used to enhance lifetime of sensor network. The dynamic routing strategy is used known as Maximum Capacity Path. The decision parameter is a residual energy of sensor node and sink is relocated for better utilization of nodes energy in the network. The strategy is known as Energy Aware Sink Re-Location where energy aware transmission range adjustment is also applied. Transmission range is dependent on health of battery used in sensor. There are three types of battery health states. These states are:-

- (i)  $0 \leq r(u) < B/3$  nodes follow  $\gamma/4$  transmission range,
- (ii)  $B/3 \leq r(u) < B/2$  nodes follow  $\gamma/2$  transmission range and
- (iii)  $B/2 \leq r(u) < B$  nodes follow  $\gamma$  transmission range

where B is battery energy,  $\gamma$  is initial transmission range, and  $r(u)$  is current residual energy of the sensor node u. There are two steps in sink relocation mechanism, first is to observe weather the relocation is needed or not and second the direction in which the sink node moves. The relocation condition might be Maximum capacity path value of each sensor nodes. This value may be below B/2 or the average residual energy of neighbor set drops below B/2. The possible relocation direction for sink node is depicted in **figure 1**. Dynamic relocation keeps energy from draining out quickly and it is better developed in this protocol but the emphasis is on residual energy only. Sink relocation could be more better approach with optimized location searching.

In the paper [4] by Sandeep Kulkarni et al. multi-hop network reprogramming protocol (MNP) technique is discussed. This is Energy Efficient Multi hop Reprogramming service, designed for sensor network. The efficient sender is selected through greedy approach. A pipeline is used for fast data propagation together with various sleep schedules. These schedules are contention sleep, no request sleep and optional initial sleep which reduce the energy consumption. The purpose of initial sleep is to reduce idle listening in initial phase of reprogramming which ultimately reduces the energy consumption. Nodes do not keep its radio ON all the time but it can take short naps, wake up and check the channel from time to time before a node gets the entire program. It saves the energy of nodes. Author observed that the selection of long contention sleep period is good for dense network because most of the collision is avoided. Short contention sleep period helps packet to reach the destination very quickly in sparse network. Noreq sleep reduces energy consumption at the end of reprogramming. Distributed approach is used for the sender selection and it causes delay during packet transmission. Network traffic is increased due to control packets overhead. For the dense

network collision is too much, long contention sleep is used to avoid collision. Long contention sleep increases latency.

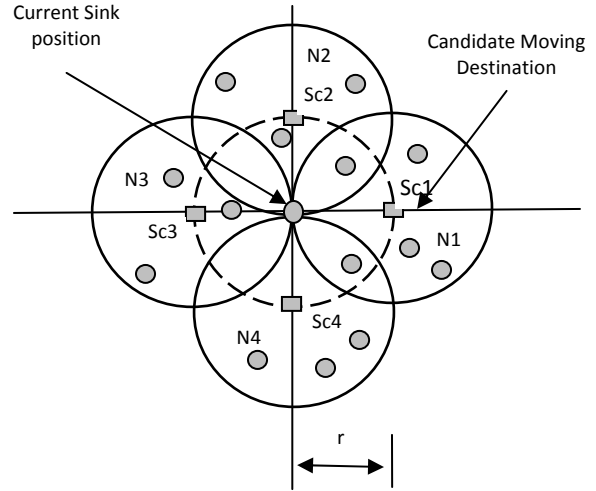


Fig.1. An illustration of the four candidate moving destination for Sink relocation

In the paper by Imad S. AlShawi, et al. [5], which resolve the problem of uneven use of network energy. A new strategy is employed where an A-star algorithm and fuzzy approach both are used simultaneously. The optimal path from source to destination is selected on the basis nodes having highest residual battery power, minimum number of hops and minimum traffic load. Fuzzy approach is easy to implement, robust and ability to approximate the nonlinear mapping. Fuzzy set is used to analyze information. Fuzzy sets allow an object to be a partial member of set. A fuzzy set A is defined by the set of ordered pairs:

$$A = \{(x, \mu_A(x)/x \in X)\}$$

Where the function  $\mu_A(x)$  is called a membership function of the object x in A, and x belongs to domain X.

A-Star algorithm is highly used as graphic search algorithm. It is combination of both depth first search and breadth first search algorithm. The evolution function is used by an A-Star path searching algorithm from source to destination is.

$$(f(n) = g(n) + h(n))$$

Where an actual cost from source node to node n is  $g(n)$ ,  $h(n)$  is an estimated cost from node n to destination node. A-Star algorithm keeps two lists for nodes which are evaluated known as OPEN and non-evaluated known as CLOSE. Best possible nodes are keeping in OPEN priority list. On the completion of an algorithm the best ever path is found out if exist and it is depend upon cost provided. The fuzzy approach is accounted for Residual energy and traffic load of node n to calculate optimal cost for node n as shown in **figure 2**. The cost generated through this fuzzy approach is  $(f(n) = NC(n) + (\frac{1}{MH(n)}))$  where MH(n) is short distance from node n to the base station and NC(n) is cost for node n taking value from [0.....1]. Author observed that A-star algorithm is better than other optimization algorithms like Genetic algorithm, Warshall Algorithm and AOD Vjr Algorithm. Another Evolutionary technique known as

Genetic algorithm is used in the work done by Sherin M. Youssef et al.[6]. They used problem specific genetic operators for improvement in computing efficiency. Distributed sensor query based application is optimized to reduce the redundancy in the network which ultimately saves the nodes energy and improve the lifetime of the network. Proposed method achieved three goals, first the set of selected nodes in the sensing region should cover entire geographical region of query, second goal is assured that all nodes are connected, and third is query processing which should be energy aware. They evaluated energy consumption of selected cover using the following equation and it can be used for fitness of chromosome  $CH_i$ :

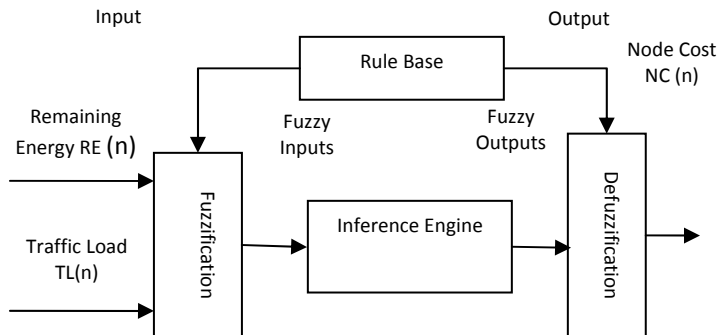


Fig 2: Fuzzy structure with two inputs

$$E_n = \sum_{i=1}^{C_s} en_i$$

Where, the consumed energy is  $en_i$  of a sensor node  $s_i$  in the  $Q$  cover chromosome, and  $C_s$  is the cover size. The process is simple selecting 50% chromosome from population  $P_t$  at time  $t$ , after which 30% population is selected from remaining 50% chromosome using crossover process and 20% population is selected through mutation from the remaining 30% chromosome.

The main contribution of the presented paper is to improve lifetime of the network by reducing traffic in the network. Proposed lifetime enhancement algorithm based on Particle Swarm Optimization technique which select the nodes for transmission to the base station. In the proposed work, improving lifetime of the network is major concern. This objective is achieved through energy saving. Various parameters are responsible for energy draining such as Traffic Load, Non uniform nodes energy usage, non effective signal usage etc. For the improvement of lifetime of the network, these parameters are chosen with their amount of participation in the network. Sum of the weighted parameters of the nodes is an objective function of proposed work and it is maximized through optimization. Nodes are selected with values (weighted sum) greater and equal to average weighted sum value of the nodes. Only selected nodes are permitted for transmission to the base station. Selection of sensor nodes for data transmission is based on Particle Swarm Optimization algorithm. It minimizes duty cycle and helps to reduce energy consumption with Sleep Wake-Up process. Count of these selected nodes is our source nodes count. Using this count the objective function is initialized.

Population of PSO is initialized with the weighted sum of each node. The algorithm applied is shown in the **Figure 3**.

The weighted sum of each parameter is calculated as:

$$W_{sum} = w_{re}RE + w_{ss}SS + w_{ql}QL$$

Where  $W_{sum}$  is sum of weights calculated from various parameters, RE is residual energy, SS signal strength and QL is queue length with their respective weights  $w_{re}$ ,  $w_{ss}$ , and  $w_{ql}$ . Maximizing the Weighted sum so that less number of nodes gets qualified and we can better save the energy of remaining nodes in the cluster. Selection of nodes having better weighted sum value than optimized weighted sum by comparing their values. These nodes are representative nodes for all the nodes within the range of them.

#### A. Network Model:

In this paper, a WSN is modeled as a collection  $N$  sensor nodes and a base station located at the center of the field, base station has large energy resource and rest of the nodes have limited energy. Sensor nodes are randomly distributed over the field. Nodes are coordinated through coordinator nodes (base station).

#### B. Energy Model:

The first order radio energy consumption model[10]-[12] is used for the nodes where  $E_{Tx}(l, d)$  is transmission energy required to send  $l$  bits data at the distance  $d$  and  $d_0$  is threshold distance for data transmission,  $E_{Rx}$  is receiving energy,  $E_{elec}$  is energy dissipated per bit to run the transmitter or receiver circuit. Transmitter amplifier energies are shown by  $E_{Fs}$  and  $E_{Tr}$ . The model is shown below:

$$E_{Tx}(l, d) = \begin{cases} l \times E_{elec} + l \times E_{Fs} \times d^2, & \text{if } d < d_0 \\ l \times E_{elec} + l \times E_{Tr} \times d^4, & \text{if } d > d_0 \end{cases}$$

### III. PROPOSED OPTIMIZATION METHOD:

Optimization can be applied for any problem related to maximization or minimization of the objective function. There are many optimization methods explained in the chapter Modern Optimization Technique of [7] for example Simulated Annealing Algorithm, Tabu Search Algorithm, Genetic algorithm, Particle Swarm Optimization and Minimum norm theorem. Among these techniques PSO is chosen for this work. PSO is bio-inspired algorithm based on movement pattern and behavior of bird folk. This method avoid converges quickly to generate result and stick to the local minima. This process is clearly defined in PSO topic.

#### A. Particle Swarm Optimization:

This algorithm is first proposed in paper [8] by J. Kennedy and R. Eberhart. This is biological inspired algorithm deals with the movement and behavioral pattern of bird folk. This pattern is investigated and observed that birds reach towards the crowded folk of birds. This crowded place can be called as an optimized place. It is confirmed for almost all herds of animals in land and water resources. In this paper, optimization of data traffic in the network is applied and investigated. Optimization in network traffic is achieved by selecting source nodes among  $N$  sensor nodes in the cluster. PSO is used as an optimization algorithm.

Particles in PSO are possible candidate solution initialized with the values provided by each sensor nodes. These particles constitute the population of particle swarm optimization algorithm. The process is iteratively running for getting best value of particle until fixed number round. During single epoch many tasks are performed by it such as updating velocity and position of the particle, setting local best position of particle and also set the global particle best position.

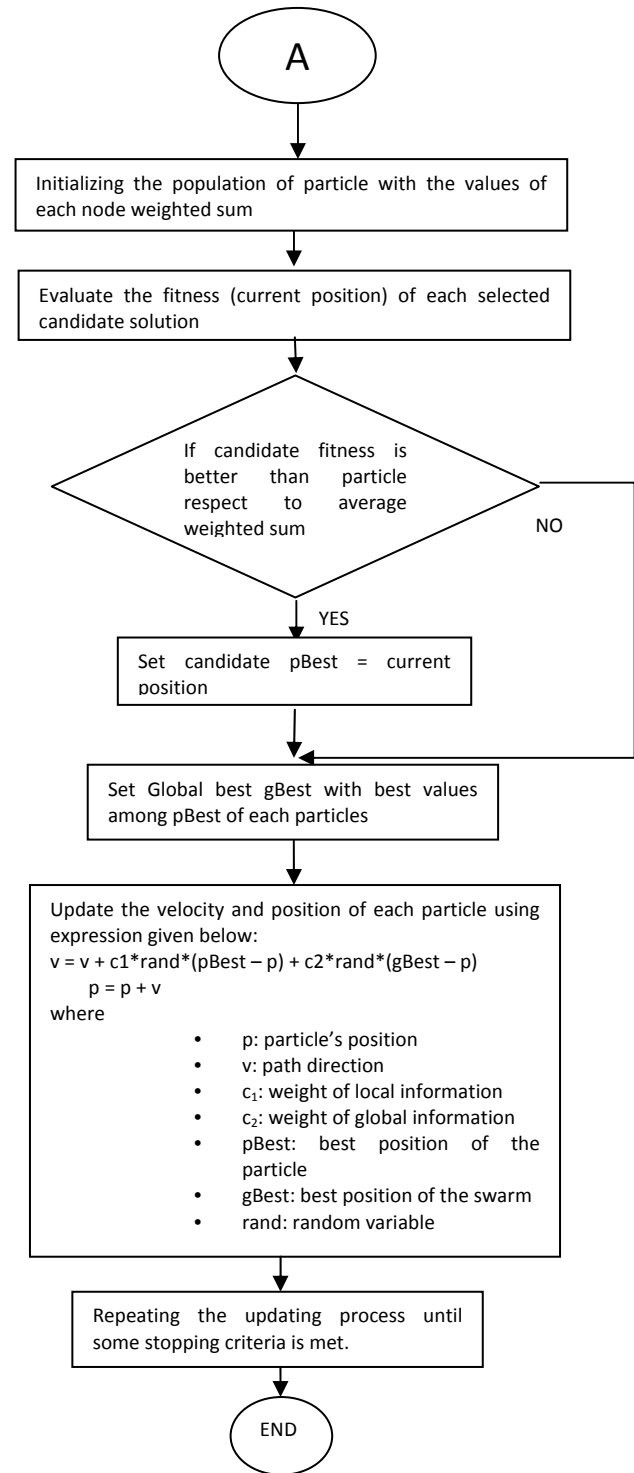
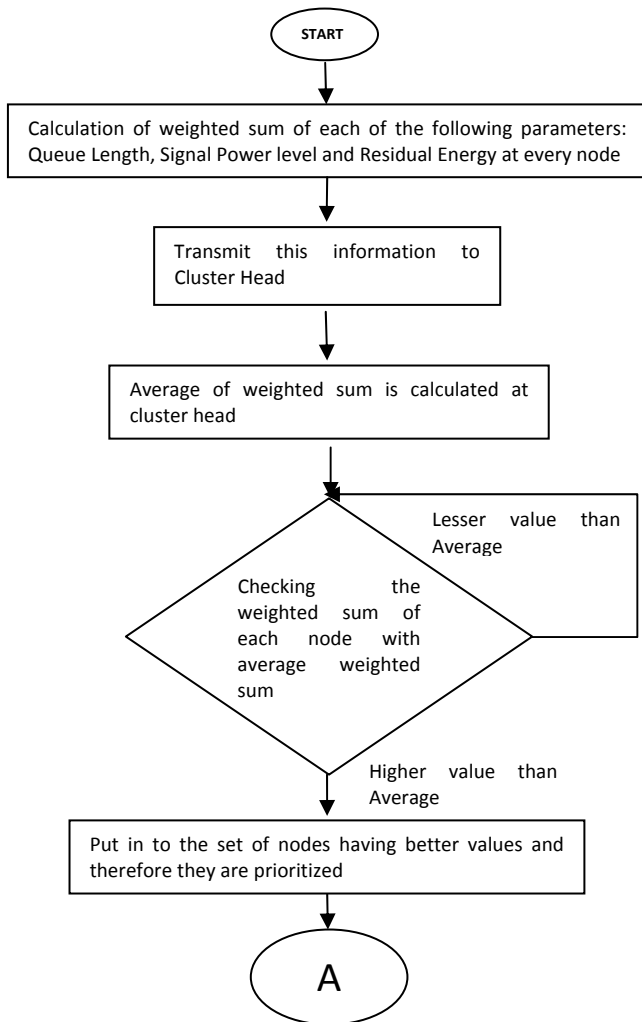


Fig 3: Algorithm PSO enabled Source selection

$$v(k+1) = v(k) + C_1 \times rand_1(k) \times (p_{Best} - p) + C_2 \times rand_2(k) \times (g_{Best} - p)$$

$$p(k+1) = p(k) + v(k+1)$$

Where  $v$  is velocity of particles,  $C_1$  and  $C_2$  are constants,  $rand$  is random numbers uniformly distributed between  $[0, 1]$ ,  $k$  is iteration count,  $p_{Best}$  is particle local best value,  $g_{Best}$  is particle global best value and  $p$  is particle value calculated so on. PSO continue iterate for the fixed number which minimized (or maximized) the solution of the problem.

#### IV. PERFORMANCE EVALUATION:

For the performance measurement the simulation is done and it is based on Network Simulator NS2.35. Algorithm is implemented using C++ and the parameters are specified through TCL script. Simulation scenario consisting of 50 sensor nodes spread around a field of 100 m<sup>2</sup>. There are many simulations have been done with varying simulation area, nodes initial energy and number of nodes. Following simulation configurations are tabulated:

TABLE 1: SIMULATION SCENARIOS

Simulation Scenario	Sim 1	Sim 2	Sim 3
Simulation area(m <sup>2</sup> )	100	50,100,150, 200	100
Transmission Range(m)	100	100	100
Initial nodes Energy (J)	1,2,3,4,5	2	2
Number of nodes	50	50	10,20,30,40,50

Simulation is done in the environment of sensor nodes sending information to the coordinator node, it contains large source of energy. All the nodes sense information for a single application such as temperature, humidity, wind speed etc. The information is disseminated to coordinator node in their TDMA schedule. The information is energy consuming because of their duplicates copies reached to destination. According to work of Dr. Wendi Hinzelman in her dissertation the energy required for transmission is much greater than energy for processing as calculated in equation.

$$E_{Tx}(l, d) = E_{Tx-elec}(l) + E_{Tx-amp}(l, d)$$

Where  $E_{Tx}(l, d)$  is transmission energy required to send  $l$  bits of data at distance  $d$  from the sensor node itself. Whereas energy for data transmission is dependent on wave propagation model which is indirectly depend upon distance. There are two propagation models:

$$E_{Tx}(l, d) = \begin{cases} lE_{elec} + l \epsilon_{friss-amp} d^2 & : d < d_{crossover} \\ lE_{elec} + l \epsilon_{two-ray-amp} d^4 & : d \geq d_{crossover} \end{cases}$$

For receiving  $l$  bits information the following equation is used:

$$E_{Rx}(l) = E_{Rx-elec}(l)E_{Rx}(l) = lE_{elec}$$

Where  $E_{Rx}(l)$  is receiving energy of  $l$  bits information at the receiver's end. The parameters  $\epsilon_{friss-amp}$  and  $\epsilon_{two-ray-amp}$  will depend upon required receiver sensitivity and receiver's noise figure. The  $d_{crossover}$  distance is used to apply radio model for power estimation required to data transmission. The transmit power is also a function of receiver threshold as :

$$P_t = \begin{cases} \alpha_1 P_{r-thresh} d^2 & : d < d_{crossover} \\ \alpha_2 P_{r-thresh} d^4 & : d \geq d_{crossover} \end{cases}$$

$$\text{Where } \alpha_1 = \frac{(4\pi)^2}{G_t G_r \lambda^2} \text{ and } \alpha_2 = \frac{1}{G_t G_r h_t^2 h_r^2}$$

$P_{r-thresh}$  is receiver threshold and can be determined noise at the receiver.

Analyzing the results in various simulations mentioned in the table bought in single conclusion i.e. a lot of energy saved and lifetime improved with the proposed algorithm.

According to first simulation, Improvement in Lifetime is tremendous as shown in **Figure 4**. This improvement is approximately 4 times of LEACH-C [9] protocol. This improvement over LEACH-C is due to less number of source nodes selected for data transmission i.e. reduced duty cycle.

This trend is further improved during increase of nodes initial energy and becomes 7 times of LEACH-C at nodes with initial energy of 5 Joules. The selection of source nodes is based on PSO optimization algorithm therefore the rate of energy draining is also reduced. There is linear growth in lifetime of network with 0.5 times of residual energy, 0.3 times of Power required and 0.2 time of queue length. Residual energy is directly proportional to lifetime of the network and larger weight to the residual energy keeps lifetime increasing linearly. Transmission power is based on distance of source node to the base station.

Two models are used for the calculation of transmitting power is discussed in Energy model, first is distance of source nodes less than threshold to the coordinator node and other is greater than it. For the nodes which are far away from the coordinator node have least possibility of selection as the source nodes and hence servility of nodes increased. It further improves the lifetime of the network.

#### A. Lifetime Improvement:

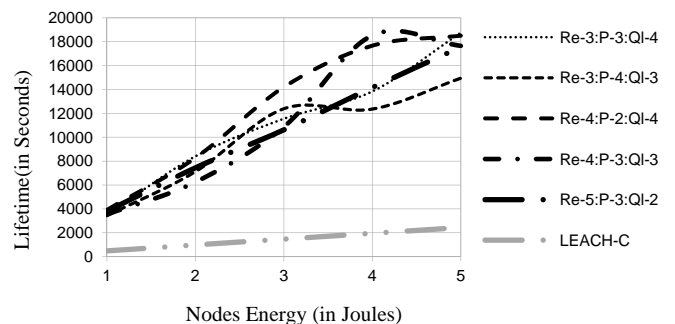


Fig 4: Lifetime of the Network VS Nodes Energy

This improvement is also approved when scenario two is applied as shown in figure 5, in the dense environment of sensor nodes lifetime improved by 7 times of LEACH-C. This improvement is easily analyzed through the graph when numbers of nodes increases from 30. Increasing nodes in the given area contributes better lifetime through their energy resource. Optimization algorithm PSO optimally choose the source nodes which are active during their TDMA schedule and rest of the nodes are in sleep state. More number of nodes in the field indicates more number of nodes in sleep state and hence more is the lifetime. This improvement is further analyzed using different weights for the given three parameters. Again Residual energy with weight 0.5 performs well as shown **Figure 5**.

Further investigation on lifetime is based on simulation areas, Nodes are placed over different sizes of field area and distance of sensor node is increased from itself to the base station.

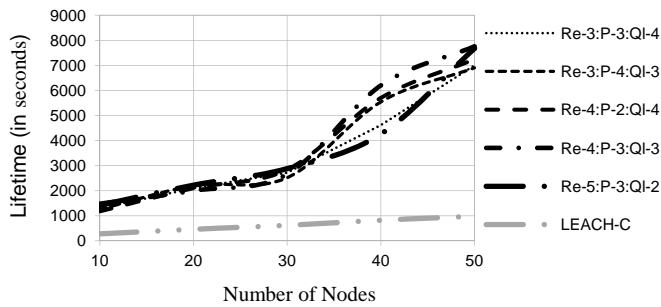


Fig 5: Lifetime of the network VS Number of Nodes

The effect of increased distance is negative for survival for sensor nodes. However, in our protocol this effect is not so vital as compared to LEACH-C. The graph shown in **Figure 6** is depicted it. Reduction of duty cycle also put its impact over here and for the line with 0.3 Residual energy, 0.4 required power and 0.3 queue length weights perform better. Field area increases therefore distance between source nodes and coordinator node is also increase. Nodes which are far away from the coordinator node will be delayed in their selection through PSO and they will survive for long time. This is clearly indicated in the graph that the slop of the improved protocol is degraded gracefully after 100 meter square. It is concluded that the nodes survival rate is improved on the larger fields.

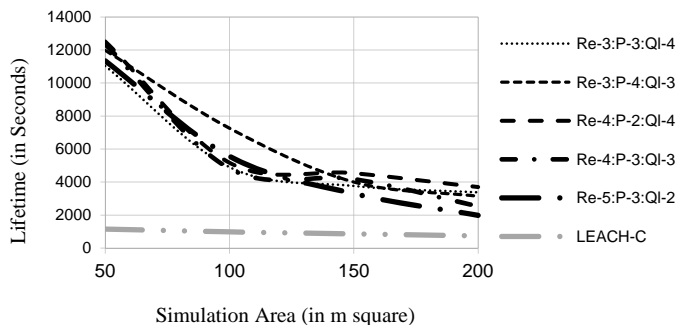


Fig 6: Lifetime Vs Simulation Area

More investigation on the protocol brought us that we saved redundant data during transmission and kept those nodes in sleep state which were sending data. The simulation result is shown in **Figures 7-8-9**

**B. Data Transmission:**

During first simulation, where about 50000 data packets are transmitted by LEACH-C protocol in simulation time and it is quarter of data packets sent through improved protocol with initial energy of 1 Joules of each node. This trend is further improved with higher initial sensor energy as we can easily analyze through the slops LEACH-C and improved protocol. Optimization algorithm helps in the selection of source node which leads to reduce the traffic in the network also. Nodes which are not selected during rounds kept their radio off until next round of selection procedure, hence nodes having greater initial energy can save their energy in larger amount. This improvement is shown in the **Figure 7**. In this figure, the reduction in data is about 5 times at nodes initial energy of 5 joules as compared to nodes energy with 1 joule. For the nodes having greater initial energy can gives better response through PSO.

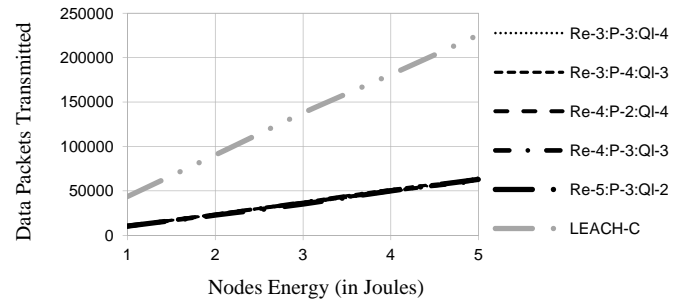


Fig 7: Data Transmission Vs Nodes Energy

In the second simulation, it is worth full to improve lifetime of the sensor network because lots of money involved in the deployment of sensor nodes in the field. For the dense environment, this improvement can be easily identified through the slops of two caterpillars in the **Figure 8**.

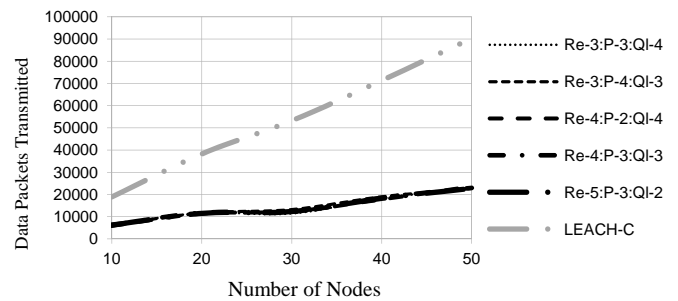


Fig 8: Data Transmission with number of nodes

Similarly PSO improves the result here also because through optimization we reduce the number of source nodes. Increase in number of nodes on the field will not affect the traffic too much.

Those nodes which are not participating in the data transmission keep into sleep state and save energy which ultimately leads to longer lifetime of the network. More number of nodes on the field more number of nodes in sleep state. In the **Figure 9**, the size of field affects the data transmission negatively but it is not consider as reduction in redundant data. When the density is low, the number of active nodes per unit area is less. Thus, fewer amounts of data are transmitted to the coordinator node.

This is due to the larger distance of some nodes to the coordinator node for efficiently data transmission. However, the difference between LEACH-C and improved protocol is clearly visible in the graph. The redundant data is minimized up-to half of its value on the field of 50 m<sup>2</sup> and it is further improved for larger fields as shown in **Figure 9**.

In the above three simulations, we have analyzed results in two important requirements of wireless sensor network which are Data and Lifetime of the network. Simulation is done through varying the three parameters like initial nodes energy, number of nodes and simulation field area. In addition to this discussion, we have analyzed some more results on the basis of parameters such as Residual energy of nodes, Minimum Power requirement for data transmission (link quality) and Queue length (Traffic Load).

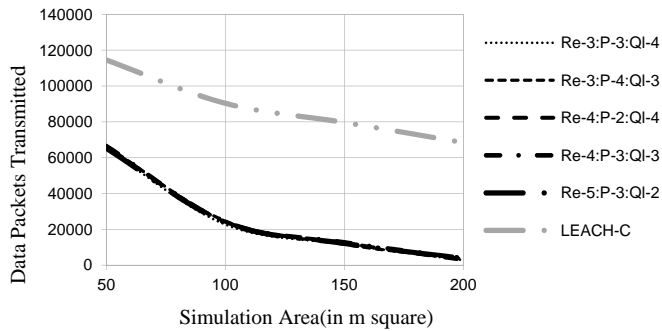


Fig 9: Data Transmission Vs simulation area

## V. CONCLUSION:

Optimization algorithm improves the technologies for better outcome and it is approved in this work. Particle swarm optimization technique reduces the redundant traffic in the network by 4 to 7 times and increase the Lifetime of the network with same proportion. Also proper weights for different parameters reached us to betterment of the work. Minimizing duty cycle using PSO saves much energy required in data transmission. Further analysis and improvement requires in the selection of parameters and their weights. Some limitation of this work is that it is applicable for single application only. With the multiple applications there is requirement of one more parameter, which is importance of information. We need to optimize this parameter also so that required information can reached to the destination timely and with less redundant data in the network.

## VI. REFERENCE:

- [1] Giuseppe Anastasi, Marco Conti, Mario Di Francesco, and Andrea Passarella, "Energy conservation in wireless sensor networks: A survey" Contents lists available at *Science Direct Ad Hoc Networks* 7 (2009) 537–568
- [2] Rashmi Ranjan Rout, Student Member, IEEE, and Soumya K. Ghosh, Member, IEEE "Enhancement of Lifetime using Duty Cycle and Network Coding in Wireless Sensor Networks" *IEEE Transaction on wireless communications*, vol. 12, NO. 2, February 2013
- [3] Chu-Fu Wang, Jau-Der Shih, Bo-Han Pan, and Tin-Yu Wu "A Network Lifetime Enhancement Method for Sink Relocation and Its Analysis in Wireless Sensor Networks" *IEEE sensors journal*, vol. 14, no. 6, June 2014
- [4] Sandeep Kilkarni and Limin Wang "Energy-Efficient Multihop Reprogramming For Sensor Networks" *ACM Transactions on Sensor Networks*, Vol. 5, No. 2, Article 16, Publication date: March 2009
- [5] Imad S. AlShawi, Lianshan Yan, Wei Pan and Bin Luo "Lifetime Enhancement in Wireless Sensor Networks Using Fuzzy Approach and A-Star Algorithm" *IEEE Sensors journal*, vol. 12, no. 10, October 2012
- [6] Sherin M. Youssef , Meer A. Hamza, Salma F. Fayed "EQOWSN: Evolutionary-based query optimization over self-organized wireless sensor networks" Available online at [www.sciencedirect.com](http://www.sciencedirect.com) *Expert Systems with Applications* 36 (2009) 81–92
- [7] S.A. Soliman and A.H. Mantawy , " Modern Optimization Techniques" with Applications in Electric Power Systems, Energy Systems, DOI 10.1007/978-1-4614-1752-1\_2, Springer Science + Business Media, LLC 2012
- [8] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. IEEE Int. Conf. Neural Netw.*, 27 Nov.–1 Dec., 1995, vol. 4, pp. 1942–1948.
- [9] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan. "An Application-Specific Protocol Architecture for Wireless Microsensor Networks". *IEEE Transaction on Wireless Communications*, vol. 1, no. 4, October 2002.
- [10] G. S. Sara and D. Sridharan, "Routing in mobile wireless sensor network: A survey," *Telecommun. Syst.*, Aug. 2013.
- [11] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayiric, "Wireless sensor networks: A survey," *Comput. Netw.*, vol. 38, no. 4, pp. 393–422, Mar. 2002.
- [12] P. Ferrari, A. Flammini, D. Marioli, and A. Taroni, "IEEE802.11 sensor networking," *IEEE Trans. Instrum. Meas.*, vol. 55, no. 2, pp. 615–619, Apr. 2006.

**Dhanpratap Singh:** Received the B.E degree from MIT Ujjain India, M.E degree from SGSITS Indore India in 2014, Perusing PhD from MANIT Bhopal India in Computer Science and Engineering discipline. He joined LNCT group of Institution Bhopal as a faculty of Information Technology department.

**Dr. Jyoti Singhai** is Professor in ECE department of MANIT Bhopal India. She has Published 40 papers in International journals, 06 papers in national journals and 72 papers in various International and national conferences after 2006. Her research area is routing protocols in Sensor networks and wireless communication networks

# Face Liveness Detection – A Comprehensive Survey Based on Dynamic and Static Techniques

Aziz Alotaibi

University of Bridgeport, CT 06604, USA

Ausif Mahmood

University of Bridgeport, CT 06604, USA

**Abstract** With the wide acceptance of online systems, the desire for accurate biometric authentication based on face recognition has increased. One of the fundamental limitations of existing systems is their vulnerability to false verification via a picture or video of the person. Thus, face liveness detection before face authentication can be performed is of vital importance. Many new algorithms and techniques for liveness detection are being developed. This paper presents a comprehensive survey of the most recent approaches and their comparison to each other. Even though some systems use hardware-based liveness detection, we focus on the software-based approaches, in particular, the important algorithms that allow for an accurate liveness detection in real-time. This paper also serves as a tutorial on some of the important, recent algorithms in this field. Although a recent paper achieved an accuracy of over 98% on the liveness NUAA benchmark, we believe that this can be further improved through incorporation of deep learning.

**Index Terms**—Face Recognition, Liveness Detection, Biometric Authentication System, Face Anti-Spoofing Attack.

## I. INTRODUCTION

Biometric authentication is an automated method that identifies or verifies a person's identity based on his/her physiological and/or behavior characteristics or traits. The Biometric authentication method is favored over traditional credential (username / password) for three reasons: first, the user must be physically present in front of the sensor for it to acquire the data. Second, the user does not need to memorize login credentials. Third, the user is free from carrying any identification such as an access token. An additional advantage of biometric systems is that they are less susceptible to Brute Force attacks. Biometric authentication can be based on physiological and/or behavior characteristics of an individual. Physiological characteristics may include, iris, palm print, face, hand geometry, odor, fingerprint, and retina etc.. Behavior characteristics are related to a user's behavior: e.g., typing rhythm, voice, and gait.

The Ideal biometric characteristics to use in a particular authentication should have five qualities[1]: robustness, distinctiveness, availability, accessibility and acceptability. Robustness refers to the lack of change of a user characteristic over time. Distinctiveness refers to a variation of the data over the population so that an individual can be uniquely identified. Availability indicates that all users possess this trait.

Accessibility refers to the ease in acquiring the characteristic using electronic sensors. Acceptability refers to the acceptance of collecting characteristic from the user. The features that provide these five attributes are then used in a biometric authentication or verification system. Verification is defined as matching of an individual's information to stored identity, whereas identification refers to whether an incoming user's data matches to any user in the stored dataset. Prior to authentication (verification or Identification), an enrollment of allowed individuals is required.

In the Enrollment mode, the users are instructed to show their behavior/physiological characteristics to the sensor. This characteristic data is acquired and passed through one of used algorithms that checks whether the acquired data is real or fake. Moreover, it ensures the quality of the image. The next step is to register the acquired data by performing localization and alignment. The acquired data is processed into a template that is a collection numbers that is stored into the database.

In the authentication phase, the biometric system includes four steps before making the final decision: Data Acquisition, Preprocessing, Feature Extraction, and Classification [2] [3].

- 1) Data acquisition: it is a sensor, such as fingerprint sensor and web camera, which captures the biometrics data with three different qualities: low, normal, and high quality.
- 2) Preprocessing: its duty is to reduce data variation in order to produce a consistent set of data by applying noise filter, smoothing filter, or normalization techniques.
- 3) Feature extraction: it extracts the relevant information from the acquired data before classifying it.
- 4) Classification: it is a method that uses the extracted features as input and assigns it to one of the output labels.

The verification mode extracts the relevant information and passes it to the classifier to compare the captured acquired data with template stored into the database to determine the match[2]. In the identification mode, the acquired data is compared with all users' template in the database to the user [3] [4]. Fig. 1 is a simple description of these three modes.

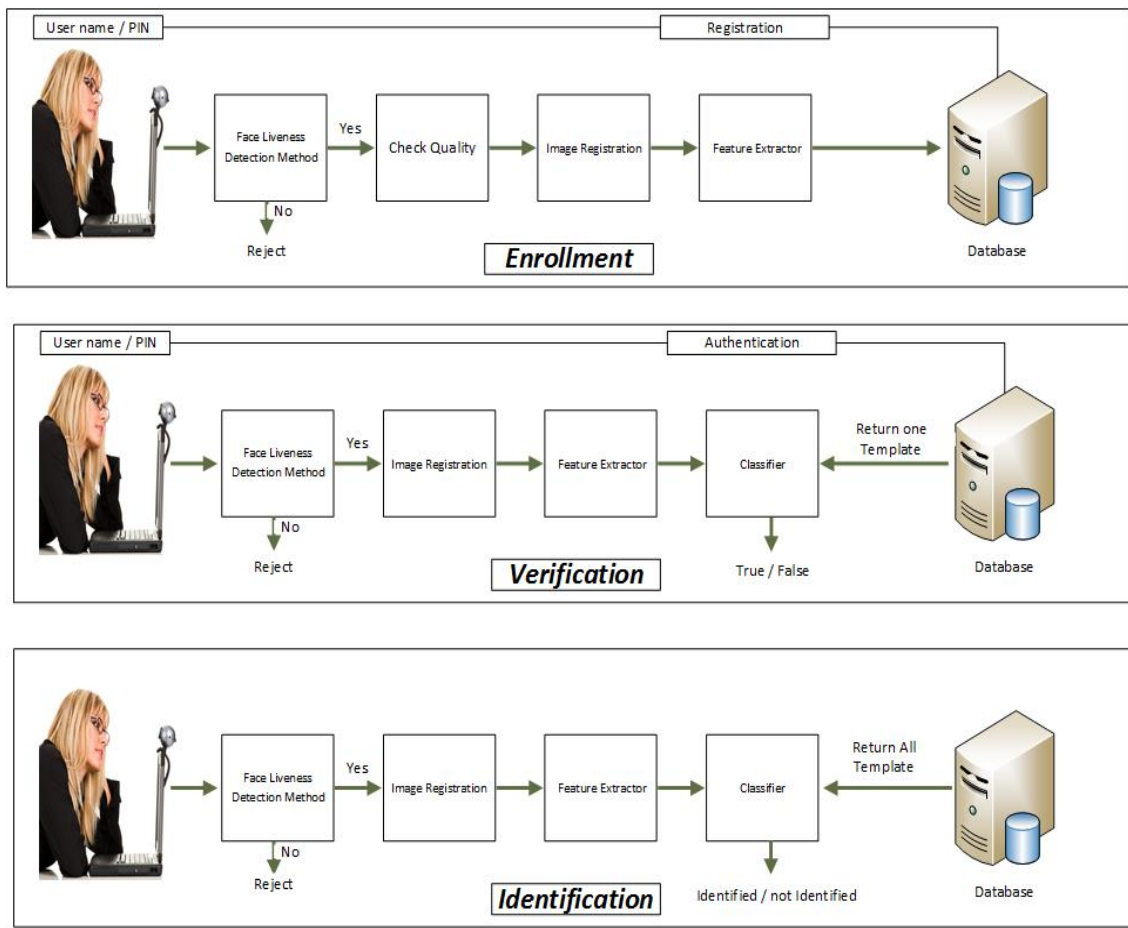


Fig. 1. Face Recognition System

For biometric systems based on face recognition, adding a face liveness detection layer to the face recognition system prevents the spoofing attacks. Before proceeding to recognize or verify the user, the face liveness checking will eliminate the possibility that a picture of the person is presented to the camera instead of the person him/herself.

The rest of the paper is organized as follows: we give a brief overview of biometric anti-spoofing method types in section III. Static and dynamic techniques are described in section IV. The experimental results and the analysis on the spoofing datasets and performance of the implemented techniques are provided in section V. Finally, we conclude this study and discuss future work in section VI.

## II. BIOMETRIC ANTI-SPOOFING METHODS

Recently the performance of the face recognition system has been enhanced significantly because of improvements found within hardware and software techniques in the computer vision field [5]. However, face recognition is still vulnerable to several attacks such as spoofing attacks. Spoofing attack techniques are getting more complex and hard to identify, especially with the advancement in

computer technologies such as 3D printers. Therefore, researchers have proposed and analyzed several approaches to protect the face recognition systems against these vulnerabilities. Based on the proposed techniques, face anti-spoofing methods are grouped into two main categories: hardware-based technique and software-based technique. First, The hardware-based technique requires an extra device to detect a particular biometric trait such as finger sweat, blood pressure, facial thermogram, or eye reflection [6]. This sensor device incorporated into the biometrics authentication system that requires the user's cooperation to detect the signal of the living body. Some auxiliary devices, such as infrared equipment, achieve higher accuracy when compared to simple devices. However, auxiliary devices are expensive and difficult to implement [7]. Second, the software-based technique extracts the feature of the biometric traits through a standard sensor to distinguish the real traits from the fake traits. The feature extraction occurs after the biometric traits are acquired by the sensor such as the texture features in the facial image [8]. The software-based techniques treat the acquired 3D and 2D traits both as 2D to extract the information feature. Therefore, the depth information is utilized to differentiate between 3D live face and flat 2D fake face images [9]. This paper covers only the software-based



techniques that can be categorized further into static-based techniques and dynamic-based techniques as described in the following section.

### III. SOFTWARE-BASED TECHNIQUES

Static-based and dynamic-based techniques are less expensive and easy to implement compared to the hardware-based technique. First, the static techniques are based on the analysis of a 2D single static image. It is non-intrusive interaction which is convenient for many users. On other hand, the dynamic techniques exploit the temporal and spatial features using a sequence of input frames. Some of the dynamic methods are intrusive interactions which force the user to follow specific instructions.

#### Static techniques:

A variety of proposed methods are presented to address the spoofing attack problems that utilize a single static image. The static-based techniques are divided into two categories: texture analysis methods and Fourier Spectrum methods:

(i) *Texture analysis methods*: these methods extract the texture properties of the facial image based on the feature descriptor. Maatta *et al.* [10] analyzed the texture of the 2D facial image using multi-scale local binary pattern (LBP) to detect face liveness. The authors applied multi- LBP operators on the 2D face image to generate a concatenated feature histogram. The histogram is fed into the Support Vector Machine (SVM) classifier in order to determine whether the facial image is real or fake. The Local Binary Pattern (LBP), introduced by Ojala *et al.* [11] is a nonparametric method that extracts the texture properties of the 2D facial image with features based on the local neighborhood [12] as shown in Figure 3. The basic LBP pattern operator for each pixel in the facial image is calculated by using the circular neighborhood as shown in Figure 2.

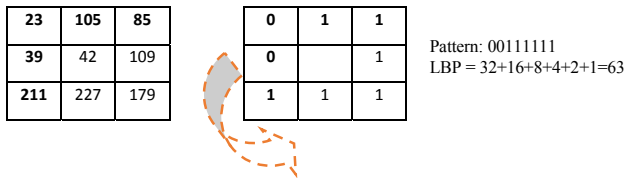


Figure. 2. The basic LBP Operator

The intensity of the centered pixel is compared with the intensity value of the pixels located within its LBP<sub>3\*3</sub> neighborhood.

$$LBP_{Point,Radius}(x_c + y_c) = \sum_{p=0}^{P-1} S(i_p - i_c)2^p$$

Where

- $x_c, y_c$  represent the center pixel

- $p$  represents the surrounding pixel
- $s(z) = \begin{cases} 1, & \text{if } z \geq 0 \\ 0, & \text{if } z < 0 \end{cases}$

Then, the center pixel will be updated with the new pixel value of 63. The LBP uses a uniform pattern to describe the texture image. If the generated binary number contains at most two bitwise 0 -1 or vice versa, then LBP is called uniform. For instance, (01111110), (1100 0000), and (0001 1000) are uniform, whereas (0101 000), (0001 0010), and (0100 0100) are non-uniform. There are 58 uniform LBP Patterns and 198 non-uniform LBP patterns. Authors applied three multi-scale LBP operators on the normalized face images: LBP<sub>8,1</sub><sup>u2</sup>, LBP<sub>8,2</sub><sup>u2</sup>, and LBP<sub>16,2</sub><sup>u2</sup>.

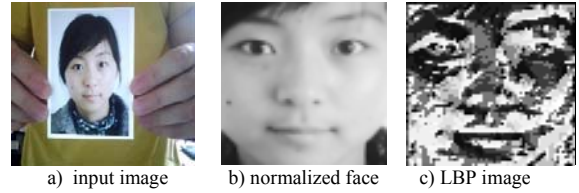


Figure. 3. Applying LBP operator on normalized face image.

The LBP<sub>8,1</sub><sup>u2</sup> was applied on a nine-block region of the normlized face, and therefore, generated uniform patterns with a 59 -bin histogram from each region . the entire image equaled a single 531-bin histogram.

The LBP<sub>8,2</sub><sup>u2</sup>, and LBP<sub>16,2</sub><sup>u2</sup> operators generates 59-bin and 243-bin histogram, respectively. The length of the concatenated feature histogram is 833. The concatenated histogram is passed through a nonlinear SVM classifier to determine whether the input face image is present or not. However, the basic LBP operator is not the only operator applied to extract the information features, other LBP variations might be used too such as transitional (tLBP), direction-coded (dLBP) and modified (mLBP). In [13], Chingovska *et al.* introduced Replay-Attack Database and studied the effectiveness of the local Binary Pattern on three types of attacks: printed photographs, photos, and videos display.



Figure.4. A frame of short videos from Replay Attack database.

The authors applied different LBP operators and studied the performance evaluation of the anti-spoofing algorithm. The study included tLBP, dLBP and mLBP. The tLBP operator is composed by comparing the two consecutive pixels value with their neighbors in a clockwise direction for all pixels apart from the central pixel value as shown in Figure 5.

$$LBP_{p,R}(x_c + y_c) = S(i_0 - i_{p-1}) + \sum_{p=0}^{P-1} S(i_p - i_{p-1})2^p$$

A direction-coded LBP operator is composed by comparing the intensity variation along the four base directions into two bits through the central pixel.

- Let's assume the original LBP<sub>p,R</sub> has P=2P' neighbors.

$$dLBP_{p,R} = \sum_{p'=0}^{P'-1} (S(i_{p'} - i_c)(i_{p'} + p' - i_c)2^{2p'} + S(|i_{p'} - i_c| - |i_{p'} + p' - i_c|2^{2p'+1}))$$

The dLBP compares the intensity of each pixel value of neighbors with the average of the intensity value in a 3 \* 3 neighborhood.

$$LBP_{Point,Radius}(x_c + y_c) = \sum_{p=0}^{P-1} S(i_p - i_c)2^p$$

Where

- x<sub>c</sub>,y<sub>c</sub> represent the center pixel
- p represents the surrounding pixel
- $s(z) = \begin{cases} 1, & \text{if } z \geq Ave \\ 0, & \text{if } z < Ave \end{cases}$

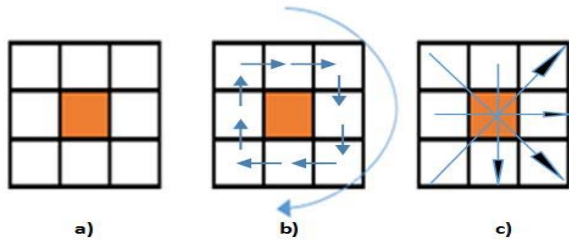


Figure. 5. a) Modified b) Transition c) Direction LBP

After applying the LBP Operators on the facial images, histograms are obtained as feature vectors. Then the applied classifier extracts the feature and determines whether the facial image is real or fake. Both linear and non-linear classifiers were examined such as Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM). The authors conducted an experiment to compare  $\chi^2$  statistics methods to other complex classifiers.

Table 1. HTER (%) OF THE CLASSIFICATION ON DIFFERENT DATABASE

	REPLAY-ATTACK	NUAA	CASIA-FASD
LBP <sub>3*3</sub> <sup>u2</sup> + $\chi^2$	34.01	-	-
LBP <sub>3*3</sub> <sup>u2</sup> + LDA	17.17	18.32	21.01
LBP <sub>3*3</sub> <sup>u2</sup> + SVM	15.16	19.03	18.17
LBP + SVM	13.87	13.17	18.21

From Table 3, we observe that the LBP extracts adequate features from the single static image which assists in the classification of fake or real faces.

The performance of the multi-scale LBP is calculated using the Half Total Error Rate (HTER). HTER is defined as the half of the sum of the False Rejection Rate (FRR) and False Acceptance Rate (FAR). HTER is used to measure the

performance on both the development sets and the test sets. Both LDA and SVM show high performance on the development sets and low performance on the test sets.

$$HTER = \frac{FRR + FAR}{2}$$

Where,

FRR = FR/ NI *False Rejection, and Number of Imposter*  
FAR = FA/ NR *False Acceptance, and Number of Rea*

Table 2. HTER (%) of classification with ( $\chi^2$ ) for different LBP operators on Replay-Attack Database.

LBP <sub>3*3</sub> <sup>u2</sup>		tLBP		dLBP		mLBP	
Dev	Test	Dev	Test	Dev	Test	Dev	Test
31.24	34.01	29.37	35.35	36.71	40.26	32.29	33.68

In [14] Kim *et al.* proposed a real-time and non-intrusive method based on diffusion speed of a single image to detect face liveness. Their idea is based on the difference in the illumination characteristic of both live and fake faces. The additive operator splitting (AOS) schema is used to compute the image diffusion [15]:

$$u^{k+1} = \frac{1}{2} ((I - 2\pi A_x(u^k)^{-1} + (I - 2\pi A_y(u^k)^{-1})u^k$$

Where  $A_x$  and  $A_y$  denote the diffusion matrices computed in column wise and row wise. The AOS schema treats every coordinate axis in the same manner, and it is unconditionally stable with large time step, e.g.  $\pi = 40$ .

To compute the diffusion speed at each pixel position(x, y):

$$s(x, y) = |\log(u^0(x, y) + 1) - \log(u^L(x, y) + 1)|$$

The features are extracted using Local pattern of the diffusion speed, so-called Local Speed Pattern (LSP):

$$LSP(x, y) = \sum_{1 \leq i \leq n} 2^{i-1} LSP^i(x, y)$$

$$LSP(x, y) = \begin{cases} 1, & \text{if } s(x, y) > (x_i, y_i) \\ 0, & \text{otherwise,} \end{cases}$$

Where  $n$  represents the number of sampling pixels.  $(x, y)$  is the centered pixel, and  $(x_i, y_i)$  denotes the position neighborhood. The extracted feature are fed into the SVM classifier to determine whether the input face is real access or fake access.

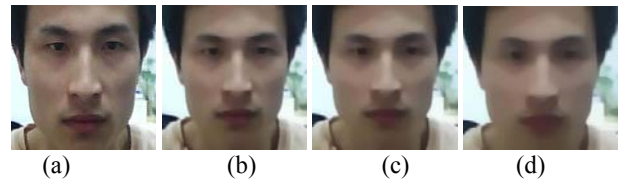


Figure 6. Example of diffusion image  $u^k$  with different iteration number and time step equals to 10. (a) original image. (b) k = 5. (c) k = 10. (d) k = 20.

Yang *et al.* [16] proposed a component-based face recognition coding approach for face liveness detection. First, the holistic face (H-Face) is divided into six components: counter, facial, left eye, right eye, mouth, and nose regions. Subsequently, counter, facial regions are further divided into 2 \* 2 grids, respectively. Moreover, the dense low-level features such (LBP, LQP, HOG, etc.) are extracted for all twelve components. Furthermore, component-based coding is performed to derive high level face representation of each one of the twelve components from low-level features. Finally, the concatenating histograms from the twelve components are fed to a SVM classifier for identification.

Table 3. Performance on NUAA, PRINT-ATTACK, and CASIA [48]

Database	Scenario	Accuracy with Metric (5)
NUAA		<b>0.977</b>
PRINT-ATTACK	Fixed(F) sub-database	<b>0.995</b>
	Hand (H)sub-database	<b>0.991</b>
	(F) and(H)sub-databases	<b>0.988</b>
CASIA	Low Quality	<b>0.987</b>
	Low Quality	<b>0.931</b>
	Warped Photo	<b>0.930</b>
	Video Photo	<b>0.997</b>
	Overall test	<b>0.898</b>

The texture analysis methods are used to extract the discriminative features for texture based classifications. However, they are less sensitive to noise in uniform regions, and their performance is degraded under the changing of lighting directions and shadowing [17].

(ii) *Methods based on Fourier spectra*: Fourier spectra is used to capture the frequency distribution of the input images to detect spoofing attacks. The structure texture of fake images are 2-D and real images are 3-D. The reflection of the light on 2D and 3D objects result in different frequency distribution. Therefore, the intensity contrast of fake images contains a less high frequency component. In [18] [19], the authors analyzed the input images using 2D Fourier spectra to extract the feature information in order to detect whether the input image is real or fake. Unlike [4][46], which used very high frequency band which is too noisy, the authors applied a Difference of Gaussian (DoG) filter that is two Gaussian filters with different standard deviation to extract the difference of the image variability. As depicted in Figure 7, DoG is applied to remove lighting variation in the input image and preserve as much features as possible without causing noise. Gaussian function with standard deviation  $\sigma_1$  as given:

$$G_{\sigma_1}(x, y) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{x^2 + y^2}{2\sigma_1^2}\right)$$

Table 4. Gaussian filter (3, 3) with  $\sigma_1=1.0$  and  $0.5$  respectively.

0.075	0.124	0.075	0.011	0.084	0.011
0.124	0.204	0.124	0.084	0.62	0.084
0.075	0.124	0.075	0.011	0.084	0.011

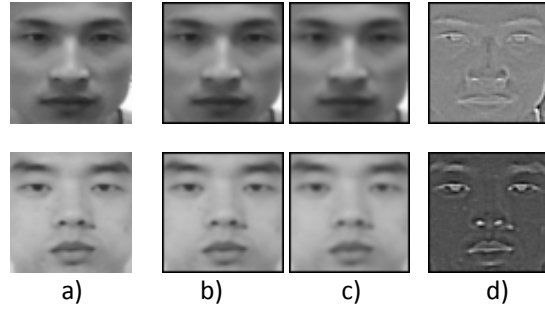


Figure. 7. (a) input image, (b)  $G_{\sigma_1}=0.5$ , (c)  $G_{\sigma_2}=0.5$ , (d) Difference of Gaussian.

Gaussian filter  $g(u, v)$  with two different standard deviations  $\sigma_1=0.5$ ,  $\sigma_2=1.0$  on the input image  $f(x, y)$  is defined as:

$$\text{DoG}(x, y) = (G_{\sigma_1}(u, v) * f(x, y)) - (G_{\sigma_2}(u, v) * f(x, y))$$

Peixoto *et al.* [19] used DoG with the Sparse Logistic Regression Model to detect the spoofing attack under extreme illumination. The sparse Logistic Regression is given as:

$$\text{Prob}(y|x) = \frac{1}{1 + \exp(-y(w^T x + b))}$$

Where  $w$  is the weight vector, and  $b$  is the intercept, And the average logistic loss is defined as:

$$\text{Loss}(w, b) = \frac{1}{m} \sum_{i=0}^m \log(1 + \exp(-y_i w^T x_i + b))$$

Since the illumination changes affect the input image, [19] used the contrast-limited adaptive histogram equalization [20] to deal with the illumination changes. In addition, Tan *et al* [18] applied the DoG and the variation Retinex-based to extract the latent reflectance features. Authors modified the sparse logistic regression to learn a “low rank” projection matrix.

Table 5. Experiment result for NUAA database [19].

Approach	Min	Mean	Max	STD
Tan <i>et al</i> “low rank” [18]	85.2%	86.6%	87.5%	0.6%
Peixoto <i>et al</i> “bad illumination” [15]	92.0%	93.2%	94.5%	0.4%

Table 4. shows that the DoG with the Sparse Logistic Regression achieved 94.5% on NUAA dataset. The result

reflects that the Fourier spectra methods have the ability to capture enough feature of the input image in order to identify the spoof attack. Further, Zang, *et al* [21] used a multiple difference of Gaussian (DoG) filters to extract the high frequency feature from the input face image. Four DoG filters are used to compute the inner and outer Gaussian variance. Let,  $\sigma_1$  represents the inner variance,  $\sigma_2$  the outer variance:

$\sigma_1 = 0.5, \sigma_2 = 1; \sigma_1 = 1.0, \sigma_2 = 1.5; \sigma_1 = 1.5, \sigma_2 = 2; \text{ and } \sigma_1 = 1, \sigma_2 = 2.$

Then the concatenated filtered images are fed into SVM classifier. Moreover, Li *et al.* [22] detected the live and fake face images based on analysis of their 2D Fourier Spectra on the face and [4] on the hair. Authors calculate the high frequency component using the high frequency descriptor equation. The high frequency descriptor of a live face should be greater than a predefined threshold  $T_{fit}$ , and the value of Fourier transform is more than the predefined threshold  $T_f$ .

$$HFD = \frac{\iint_{\Omega=\{(u,v)|\sqrt{u^2+v^2}>\frac{2}{3}f_{max} \text{ and } |F(u,v)|>T_f\}} |F(u,v)| du dv}{\iint |F(u,v)| du dv - F(0,0)} \times 1000$$

Where  $F(u, v)$  represents Fourier transform of the input image,  $f_{max}$  denotes the highest radius frequency of  $F(u, v)$ ,  $T_f$  and  $T_{fd}$  are a predefined threshold. The denominator denotes the total energy in frequency domain which is the sum of Fourier coefficients relative to direct coefficient.

*Dynamic methods:*

Dynamic methods rely on the detection of motion over the input frames sequence to extract dynamic features enabling the distinction between real face from fake face. Pereira *et al.* [23] proposed a novel countermeasure against face spoofing based on Local Binary Pattern from three Orthogonal Plans (LBP-TOP) which combines both space and time information into a multi-resolution texture descriptor. Volume Local Binary Pattern (VLBP) [24], which is an extension to the Local Binary Pattern, was introduced to extract the features from dynamic texture.

$$VLBP_{L,P,R} = \sum_{q=0}^{3P+1} f(i_c - i_q) 2^q$$

And  $f(x)$  is defined:

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$$

VLBP considers the frame sequence as parallel sequence planes, unlike LBP-TOP which considers the three orthogonal planes intersecting the pixel of the center for each pixel in a frame sequence. Orthogonal planes consist of XY plane, XT plane, and YT plane, where T represents the time

axis. Three different histograms are generated from the three orthogonal planes and then concatenated and fed to the classifier. In [25], Bharadwaj *et al* presented a new framework for face video spoofing detection using motion magnification. The Eulerian motion magnification technique is applied to enhance the facial expressions exhibited by clients in a captured video. In the feature extraction stage, the authors used both multi-scale LBP ( $LBP^{u2}_{8,1}$ ,  $LBP^{u2}_{8,2}$ , and  $LBP^{u2}_{16,2}$ ), and Histogram of Oriented Optical Flows (HOOF). The optical flow is the pattern of the apparent motion estimation technique that computes the motion of each pixel by solving the optimization problem. The PCA is used to reduce the dimensionality of HOOF vector. Finally, LDA classifier is used to classify the concatenated HOOF to detect whether the video input is real or face access.

Further, Pan *et al.* [26] proposed an Eyeblinking behavior method to detect spoofing face recognition based on an unidirectional conditional graphic framework. The eyeblinking behavior is represented as temporal image sequences after being captured. The unidirectional conditional model reduces the computational cost. It is easy to extract the feature from the intermediate observation, where the conditional model increases the complexity and makes the problem more complicated. The authors developed an eye closity method by computing discriminative information for eye states:

$$u_m(I) = \sum_{i=1}^M (\log \frac{1}{\beta_i}) h_i(I) - \frac{1}{2} \sum_{i=1}^M \log \frac{1}{\beta_i}$$

Where,

$$\beta_i = \epsilon_i / (1 - \epsilon_i)$$

And  $u(I)$  is the eye closity, and  $h_i(I) : R^{d(I)} \rightarrow \{0,1\}, i = 1, 2, \dots, M$  is a set of binary weak classifier. The input  $I$  has two states, open eye: (0) and closed eye: (1).  $\beta$  represents a closing eye state. The Adaboost algorithm is used to classify the positive value as closed eye and negative value as open eye. A blinking activity sequence of eye closity is shown in Figure 8.

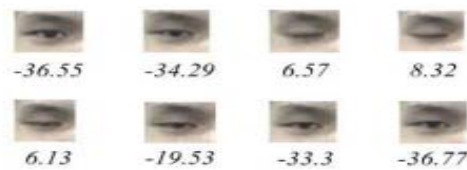


Figure 8. Illustration of the closity for a blinking activity sequence [26].

In [27] Wen *et al.* proposed a face spoof detection algorithm based on Image Distortion Analysis (IDA). Four different types of IDA features (specular reflection, blurriness, color moments, and color diversity) have been extracted from the input frame.



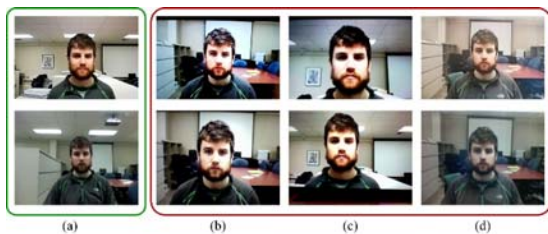


Figure 9 Example images of genuine and spoof faces. (a) Genuine faces. (b) Spoof face generated by video replay attack. (c) Spoof face generated by iPhone. (d) Spoof face generated by printed attack [27].

The IDA features are concatenated together to produce a 121-dimensional IDA feature vector. The feature vector is fed into an ensemble classifier. It is a multiple SVM classifier to distinguish between real and spoof faces. Their detection algorithm is extended to the multi-frame face detection in the playback video using a voting based schema. IDA technique is computationally expensive and consumes time in the case of using multi-frames to detect the spoofing attack.

In [28], Singh *et al.* proposed a framework to detect the face liveness using eye and mouth movement. Challenge and response are randomly generated in order to detect and calculate the eye and mouth movements using Haar Classifier. The eye openness and closeness can be measured during the time interval while the mouth is measured using the teeth Hue Saturation Value (HSV). If the calculated response is equal to number of the challenges, the proposed system will recognize the user as live.

Kim *et al.* [29] presented a new novel method for face spoofing detection using camera focusing. Two sequential images were taken with two different focusing: on nose (IN) and on ears (IE). SUM Modified Laplacian (SML) is used to measure the degree of focusing for both nose (SN) and ears (SE). After calculating SMLs, the SN is subtracted from SE to maximize the SML gap between nose and ears regions. If the sum of difference of SMLs (DoS) shows similar pattern consistently, the user is live. Otherwise it is fake. The difference in the patterns can be used as features to detect the face liveness.

In [30], Kim *et al.* segmented the video input into the foreground and background regions to detect the motion and similarity in order to prevent image and video spoofing attacks. The authors used a structural similarity index measure (SSIM) to measure the similarity between the initial background region and the current background region. And the background motion index (BMI) is proposed to show the amount of motion in the background compared with foreground region. The motion and similarity in the background region should contain significant information to indicate liveness detection.

In [31], Tirunagari *et al.* used a recent developed algorithm called Dynamic Mode Decomposition (DMD) to prevent replay attacks. The DMD algorithm is a mathematical method developed to analyze and extract the relevant modes from empirical data generated by non-linear complex fluid flows. The DMD algorithm can represent the

temporal information of the entire input video as a single image with the same dimensions as those images contained in the recorded video. The authors modified the original MDM that uses QR-decomposition and used LU decomposition to make it more practical. The DMD is used to capture the dynamic visual in the input video. The feature information is extracted from the visual dynamic using the LBP and fed to SVM classifier.

Yan *et al.* [32] proposed a novel liveness detection method based on three clues in both temporal and spatial domain. First, non-rigid motion analysis is applied to find the non-rigid motion in the local face regions. The non-rigid motion can be exhibited in the real face while many fake faces cannot. Second, in face-background consistency both the fake face motion and background motion are consistent and dependent. Finally, the banding effect is the only spatial clue that can be detected in the fake images, because the image quality is degraded due to the reproduction. Their techniques show a better generalization capability on different datasets.

In [33] [34] [35] the authors analyzed the optical flow in the input image to detect the spoofing attacks. The optical flow fields generated by the movement of two-dimensional object and by three-dimensional object are utilized to distinguish between real face from fake face images. They calculate the difference in the pixel intensity of image frames to extract the motion information. The motion information are fed to the classifier to determine whether the input images are real or not.

### 3D mask:

In previous studies, 2D attacks are performed by showing printed photos or videos to the system on flat surface. However, with the advancement in 3D printing technologies, the detection of the 3D mask against 2D mask has become more complex and harder to identify [34]. Since the liveness detection and motion analysis fail to detect and protect the system against 3D mask attacks, texture analysis method is one of reliable approaches that can detect a 3D mask.



Figure 10. 3D face masks obtained from ThatsMyFace.com

In [36] [37] [38], Local Binary Pattern and its variations are proposed to protect face recognition system against 3D

mask attacks. As explained before, LBP is used to extract features and generate a histogram using 3D MAD database. The LBP histogram matching using  $x_2$  is applied to compare test samples with a reference histogram. Additionally, both linear (LDA) and non-linear (SVM) classifiers are tested. Principle Component Analysis (PCA) is used to reduce dimensionality, while 99% of the energy is preserved. The Inter Session Variability (ISV), an extension of Gaussian Mixture Models approach, is applied to estimate more reliable client models by modelling and removing within-client variation using a low-dimensional subspace [39]. Their experimental result shows that using LDA classification is more accurate in 3D mask attacks, especially in case of 3DMAD database.

#### IV. EXPERIMENTAL RESULTS ANALYSIS

In this section, we provide detailed information about the five diverse datasets that cover the following three types of attacks: printed, video records, 3D mask. Furthermore, we evaluate and compare the performance of existing algorithms on three datasets: NUAA, CASIA and REPLAY-ATTACK databases. Finally, we summarize the most popular used algorithms in static and dynamic techniques.

##### A) Anti-spoofing Datasets:

1) NUAA Photograph Imposter Database [18], which was released in 2010, is publicly available and widely used for evaluating face liveness detection. The database consists of 12,614 of both real-face and fake-face attack attempts of 15 subjects which has been collected in three sessions with about a two week interval between two sessions. For each subject in each session, the subject was asked to directly face the web camera to in order to capture a series of face images with a natural expression and no apparent movement (with 20 frame rate of 20fps).

Table 6. NUAA Database

Training Set				
	Session 1	Session 2	Session 3	Total
<b>Client</b>	889	854	0	1743
<b>Imposter</b>	855	893	0	1748
<b>Total</b>	1744	1747	0	3491
Test Set				
	Session 1	Session 2	Session 3	Total
<b>Client</b>	0	0	3362	3362
<b>Imposter</b>	0	0	5761	5761
<b>Total</b>	0	0	9123	9123

The imposter images were collected by printing the capture images on three different hard-copies: 6.8cm x 10.2 cm, 8.9

cm x 12.7 cm, and A4 paper. The database images were resized to 64 x 64 and divided into a train set with a total of 3,491 images and a test set with a total of 9,123 images. Here, the train set contains samples from the first and second session, and the test set contains only the third session. No overlapping between the train set and test set occurred.

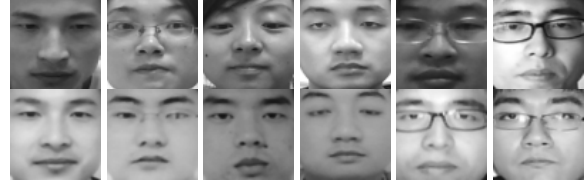


Fig. 11 Example of NUAA Database (Top: Live photo, Bottom: Fake photo)

2) Replay-Attack Database [13] consists of 1300 short videos of both real-access and spoofing attacks of 50 different subjects. Each person recorded a number of video with a resolution of 320 x 240 pixels under two different conditions: (1) the controlled condition contained a uniform background fluorescence lamp illumination; and (2) the adverse condition contained a non-uniform background and day-light illumination. The spoof attacks were generated by using one of the following scenarios: (1) print using hard copy, (2) phone using iPhone screen, and (3) tablet using iPad screen. Each spoof attack video was captured in two different attack modes: hand-based attacks and fixed support attacks [32]. The Replay-Attack database is divided into three subsets: training, development, and testing.

Table 7. Replay-Attack Database

Type	Training Fixed   hand	Development Fixed   hand	Test Fixed   hand	
<b>Genuine face</b>	60	60	80	200
<b>Print-attack</b>	30 + 30	30 + 30	40 + 40	100 + 100
<b>Phone-attack</b>	60 + 60	60 + 60	80 + 80	200 + 200
<b>Tablet-attack</b>	60 + 60	60 + 60	80 + 80	200 + 200
<b>Total</b>	360	360	480	1200

3) CASIA Face anti-Spoofing Database (FASD) [21] is publicly available, and was released in 2012. The database contains 600 short videos of both real-access and spoofing attack of 50 subjects. Each subject has 12 video clips in the database (3 real-access and 9 spoofing attacks). The genuine faces are collected with three different qualities: low quality video using USB camera, normal quality video using USB camera, and high quality video using high definition camera. The fake faces are collected using three different kind of attacks: warped photo attack, cut photo attack, and video playback attack. The database is divided into training set which contains 20 subjects, and testing set which contains 30 subjects.

4) MSU Mobile Face Spoofing Database (MFSD) [27] was released in 2014, and contains 440 videos clips consisting of 110 real access and 330 spoofing attack of 55. These videos are captured by using a Mac laptop camera with a resolution of 640 x 480, and also an Android Camera that captures videos with a resolution of 720 x 480. The video duration length is 12 second and the average of each frame is 30fps.  
5) 3D Mask Attack Database (3DMAD) [36] is the first 3D face spoofing attack publicly available. The database consists of 76500 frames that include 17 different subjects [35] recorded with a Microsoft Kinect sensor. The videos are recorded in three different sessions: The first two sessions are real-access videos and the third session is a mask attack.

*B) Performance Evaluation and analysis:*

In this subsection, we study and evaluate the effectiveness of static and dynamic techniques on the face spoofing datasets. We found that the static technique is often difficult in detecting the spoofing attacks because it uses a single static image. Many algorithms such as texture and Fourier spectra components have been introduced to solve these difficulties. We evaluate the most used static methods in face liveness detection on the NUAA database as shown in Table 8.

Table 8. Performance comparison on the NUAA Database

Methods	LTV [40]	Multi-DoG [21]
Accuracy	<b>68.44%</b> * [14]	<b>81.80%</b>
Methods	HDF [22]	DoG-Sparse [18]
Accuracy	<b>84.50%</b>	<b>87.50%</b>
Methods	MLBP [10]	DoG-Sparse Logistic [19]
Accuracy	<b>92.70%</b>	<b>94.50%</b>
Methods	CDD [16]	DS-LSP [14]
Accuracy	<b>97.70%</b>	<b>98.45%</b>

The performance of some proposed approaches on NUAA database are listed in Table 8: Logarithmic Total Variation (LTV) [40]; Multiple Difference of Gaussian (DoG-M) [21]; High Frequency Descriptor (HDF) [22]; Difference of Gaussian with Sparse Low Rank Bilinear Logistic Regression (DoG-Sparse) [18]; Multiple Local Binary Pattern (MLBP) [10]; DoG-Sparse Logistic Regression (DoG-Sparse Logistic) [19]; Component Dependent Descriptor (CDD) [16]; Diffusion Speed with Local Speed Pattern (DS-LSP) [14]. The texture analysis has proven to be successful in extracting the feature information from the single static image, since it achieves the best performance of 98.45% by using the Local Speed Pattern [14]. Although, the texture analysis is faster, and has a low computational complexity, the texture analysis has failed when it is applied on cross-database where the training and test sets are from different databases. Moreover, the texture analysis and fourier spectra algorithm are affected by image quality and brightness, which might reduce the performance.

In most dynamic experiments, both spatial and temporal features are utilized to improve the performance of algorithms. Thus, dynamic methods are slower. We only evaluated the performance of dynamic methods on the REPLAY-ATTACK and CASIA databases because they contain short videos with different attack types.

Table 9. HTER (%) OF THE CLASSIFICATION ON Replay-Attack dataset [55].

	Methods	Replay-Attack
Chingovska et al [13]	LBP + LDA	17.17
	LBP + SVM	15.16
Pereira et al [23]	LBP-TOP <sub>8,8,8,1,1[1-6]</sub> <sup>u2</sup> + SVM	11.15
	LBP-TOP <sub>8,4,4,1,1[1-6]</sub> <sup>u2</sup> + SVM	9.03
	LBP-TOP <sub>8,8,8,1,1[1-4]</sub> <sup>u2</sup> + SVM	7.95
	LBP-TOP <sub>8,8,8,1,1[1-2]</sub> <sup>u2</sup> + SVM	7.60
Tirunagari et al [31]	DMD+LBP+SVM <sup>f</sup>	3.75
	DMD +SVM <sup>f</sup>	7.50
	PCA+SVM <sup>f</sup>	21.50
	PCA+LBP+SVM	17.11

As shown in Table 10, LBP-TOP planes generate better results when compared with the basic LBP operator. The combination of both spatial and time information into a multiresolution texture show improvement from a HTER of 15.16% to 7.60%. Moreover, the non-linear classifier (SVM) shows a minor improvement over using the LDA classifier. The best performance achieved on Replay-Attack datasets is HETR of 3.75% by applying the DMD with LBP to extract the features of the visual dynamic. As Table 10 shows, using LBP features with DMD illustrates better performance than using only DMD features.

Table 10. HTER (%) OF THE CLASSIFICATION ON CASIA dataset [55].

	Methods	CASIA
Chingovska et al [13]	LBP + LDA	21.01
	LBP + SVM	18.17
Pereira et al [23]	LBP-TOP +SVM	23.75
Tirunagari et al [55]	DMD+LBP+SVM	21.75
	DMD +SVM <sup>f</sup>	29.50
	PCA+SVM <sup>f</sup>	33.50
	PCA+LBP+SVM	24.50

Table 10 shows that most of the proposed methods do not perform well on the CASIA dataset when compared with NUAA and Replay-Attack datasets. The CASIA dataset has less training samples. In addition, some of the real faces were captured with over saturated exposure making them look like spoof faces. The best performance on the CASIA dataset is a HTER of 21.01%, which is still considered to be a high error rate.

C) Face liveness detection algorithm summary

Table 11. summary of the most used face anti-spoofing methods.

Face Anti-spoofing Methods					
Year of Publication	Reference	Methods	Algorithms and methodology	Database	Attack
2004	Li et al [22]	Static	Detected the live and fake face image based on analysis of their 2D Fourier Spectra. The structure texture of the fake images are 2-D and the real images are 3-D. the reflection of the light on 2D and 3D objects result in different frequency distribution.	Private	Photo
2010	Tan et al [18]	Static	Analyzed the input images using 2D Fourier spectra to explore the feature information, and used The sparse Logistic Regression Model as classifier	Public, NUAA	Photo
2011	Peixoto et al [19]	Static	applied Difference of Gaussian (DoG) filter to remove lighting variation in the input image and preserve as much features as possible without causing noise.	Public , NUAA , Yale face Database	Photo
2011	Maatta et al [10]	Static	Analyzed the texture of the 2D facial image using multi-scale local binary pattern (LBP) to detect the face liveness.	Public, NUAA	Photo
2012	Chingovska et al [13]	Static	Applied different LBP operators and studied the performance evaluation of the anti-spoofing algorithm. The study included transitional (tLBP), direction-coded (dLBP) and modified (mLBP).	Public , NUAA, REPLAY-ATTACK , CASIA	Phtot , video
2012	Zhang et al [21]	Static	Used a multiple difference of Gaussian (DoG) filters to extract the high frequency feature from the input face image.	Public, CASIA- FASD	Video
2013	Yang et al [16]	Static	Introduced a component-based face recognition coding approach. Component-based coding is performed to derive the high level face representation of each one of the twelve components from low-level features.	Public, NUAA, CAISA-FASD, PRINT-ATTACK.	Photo, Video
2013	Erdogmus and Marcel [36]	Static – 3D	Applied the LBP to extract the feature infomation.	Public, 3D MAD	Video
2013	Kose et al [38]	Static – 3D	Applied the multi-scale LBP to extract the feature infomation.	Non-Public, Morpho	Video, Mask
2014	Erdogmus and Marcel [37]	Static – 3D	Evaluate both 3D MAD and Morpho databse using LBP based anti-spoofing methods.	Public, 3D MAD , Non-Public, Morpho	Phto , Mask
2015	Kim et al [14]	Static	Calculate diffusion speed of a single image to detect face liveness. Which is based on the difference in the illumination characteristic of both live and fake faces	Public , NUAA, REPLAY-ATTACK. Private, SFL.	Photo , Video
2007	Pan et al [26]	Dynamic	proposed an Eyebinking behavior method to detect spoofing face recognition based on an unidirectional conditional graphic framework	Public, Blinking video Database	Video
2011	Kim et al [30]	Dynamic	segmented the video input into foreground and background regions to detect the motion and similarity	Private	Video
2012	Pereira et al [23]	Dynamic	Their proposed method is based on the Local Binary Pattern from three Orthogonal Plans (LBP-TOP) which combines both axis. Three different histograms are generated from the three orthogonal planes.	Public, Replay-Attack	Video
2013	Bharadwaj et al [25]	Dynamic	Proposed a new framework for face video spoofing detection using motion magnification.	Public, PRINT-ATTACK, REPLAY-ATTACK	Phtot , Video
2013	Kim et al [29]	Dynamic	Two sequential images were taken with two different focusing: on nose (IN) and on ears (IE). SUM Modified Laplacian (SML) is used to measure the degree of focusing for both nose (SN) and ears (SE).	Private	Photo
2014	Singh et al [28]	Dynamic	proposed a framework to detect the face liveness using eye and mouth movement	Public , face94	photo ,Video
2015	Wen et al [27]	Dynamic	Used four different types of Image Distortion Analysis (IDA) feature (specular reflection, blurriness, color moments, and color diversity) have been extracted from the input frame.	Public , REPLAY-ATTACK, CASIA , MUS MSFD	video
2015	Tirunagari et al [31]	Dynamic	Used a recent developed algorithm called Dynamic Mode Decomposition (DMD). The DMD is used to capture the dynamic visual in the input video. The feature information is extracted from the visual dynamic using the LBP	Public , print-attack, replay-attack, CASIA	Photo, video



## V. CONCLUSION

Face liveness detection is an important precursor to online face recognition. We provide a comprehensive review of the techniques for liveness detection which are categorized into static and dynamic groups. Most Static techniques use either texture analysis methods such as Local Binary Pattern operators or Fourier spectra methods such as high frequency descriptor. The texture analysis is more powerful in extracting discriminative features such as MLBP operator. However, its performance degrades under the changing of lighting directions and shadowing. The Fourier spectra have the ability to capture the high and low frequency from the input face to detect the spoofing attack. Using the Difference of Gaussian with Sparse Logistic Regression has achieved 94% on NUAA dataset, where MLBP only achieved 92%. The Fourier spectra are sensitive to brightness effect, which affects the DoG to fail at detecting the border.

The dynamic techniques are based on the detection of motion over the input frames sequence to explore dynamic features to differentiate between real faces and fake faces. Since dynamic techniques utilize more than one frame, dynamic techniques achieve better performance compared with static techniques. Thus, dynamic techniques are slow and difficult to implement. Further, some of the dynamic techniques require the users to follow some instructions to validate their presence, but not all users may cooperate in this respect. This makes the dynamic methods not a favorable technique to use in the face liveness methods.

There are many different factors that might affect the performance of some of the proposed static and dynamic techniques such as media quality, illuminations and user cooperation. Some studies trained their proposed methods using low quality media, making their technique vulnerable to the use of high quality media vice versa. The best result for liveness detection achieved on the NUAA dataset was 98% using Local speed Pattern to extract the feature from diffused speed image. However, we believe that using the Deep learning neural networks can exploit and extract more complex features. The Deep learning networks have the capability to learn and to capture the discriminative and higher level features that cannot be captured by hand-crafted features such as Local Binary Pattern, and thus can lead to higher liveness detection accuracy.

## REFERENCE

1. Wayman, J., et al., *An introduction to biometric authentication systems*. 2005: Springer.
2. Jain, A.K., A. Ross, and S. Pankanti, *Biometrics: a tool for information security*. Information Forensics and Security, IEEE Transactions on, 2006. **1**(2): p. 125-143.
3. Kataria, A.N., et al. *A survey of automated biometric authentication techniques*. in *Engineering (NUiCONE), 2013 Nirma University International Conference on*. 2013.
4. Alotaibi, A. and A. Mahmmud. *Enhancing OAuth services security by an authentication service with face recognition*. in *Systems, Applications and Technology Conference (LISAT), 2015 IEEE Long Island*. 2015.
5. Weiwen, L. *Face liveness detection using analysis of Fourier spectra based on hair*. in *Wavelet Analysis and Pattern Recognition (ICWAPR), 2014 International Conference on*. 2014.
6. Galbally, J., S. Marcel, and J. Fierrez, *Biometric Antispoofing Methods: A Survey in Face Recognition*. Access, IEEE, 2014. **2**: p. 1530-1552.
7. Socolinsky, D.A. and A. Selinger. *A comparative analysis of face recognition performance with visible and thermal infrared imagery*. in *Pattern Recognition, 2002. Proceedings. 16th International Conference on*. 2002.
8. Pietikäinen, M. and A. Hadid, *Texture features in facial image analysis*, in *Advances in Biometric Person Authentication*. 2005, Springer. p. 1-8.
9. Choudhury, T., et al. *Multimodal person recognition using unconstrained audio and video*. in *Proceedings, International Conference on Audio-and Video-Based Person Authentication*. 1999. Citeseer.
10. Maatta, J., A. Hadid, and M. Pietikainen. *Face spoofing detection from single images using micro-texture analysis*. in *Biometrics (IJCB), 2011 International Joint Conference on*. 2011.
11. Ojala, T., M. Pietikainen, and T. Maenpaa, *Multiresolution gray-scale and rotation invariant texture classification with local binary patterns*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2002. **24**(7): p. 971-987.
12. Ahonen, T., A. Hadid, and M. Pietikainen, *Face Description with Local Binary Patterns: Application to Face Recognition*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2006. **28**(12): p. 2037-2041.
13. Chingovska, I., A. Anjos, and S. Marcel. *On the effectiveness of local binary patterns in face anti-spoofing*. in *Biometrics Special Interest Group (BIOSIG), 2012 BIOSIG - Proceedings of the International Conference of the*. 2012.
14. Wonjun, K., S. Sungjoo, and H. Jae-Joon, *Face Liveness Detection From a Single Image via Diffusion Speed Model*. Image Processing, IEEE Transactions on, 2015. **24**(8): p. 2456-2465.
15. Weickert, J., B.M.T.H. Romeny, and M.A. Viergever, *Efficient and reliable schemes for nonlinear diffusion filtering*. Image Processing, IEEE Transactions on, 1998. **7**(3): p. 398-410.
16. Jianwei, Y., et al. *Face liveness detection with component dependent descriptor*. in *Biometrics (ICB), 2013 International Conference on*. 2013.
17. Xiaoyang, T. and B. Triggs, *Enhanced Local Texture Feature Sets for Face Recognition Under*

- Difficult Lighting Conditions*. Image Processing, IEEE Transactions on, 2010. **19**(6): p. 1635-1650.
18. Tan, X., et al., *Face liveness detection from a single image with sparse low rank bilinear discriminative model*, in *Computer Vision–ECCV 2010*. 2010, Springer. p. 504-517.
19. Peixoto, B., C. Michelassi, and A. Rocha. *Face liveness detection under bad illumination conditions*. in *Image Processing (ICIP), 2011 18th IEEE International Conference on*. 2011.
20. Zuiderveld, K. *Contrast limited adaptive histogram equalization*. in *Graphics gems IV*. 1994. Academic Press Professional, Inc.
21. Zhiwei, Z., et al. *A face antispoofing database with diverse attacks*. in *Biometrics (ICB), 2012 5th IAPR International Conference on*. 2012.
22. Li, J., et al. *Live face detection based on the analysis of fourier spectra*. in *Defense and Security*. 2004. International Society for Optics and Photonics.
23. Pereira, T., et al. *LBP– T OP based countermeasure against face spoofing attacks*. in *Asian Conference on Computer Vision*. 2012.
24. Guoying, Z. and M. Pietikainen, *Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2007. **29**(6): p. 915-928.
25. Bharadwaj, S., et al. *Computationally Efficient Face Spoofing Detection with Motion Magnification*. in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*. 2013.
26. Gang, P., et al. *Eyeblink-based Anti-Spoofing in Face Recognition from a Generic Webcam*. in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. 2007.
27. Di, W., H. Hu, and A.K. Jain, *Face Spoof Detection With Image Distortion Analysis*. Information Forensics and Security, IEEE Transactions on, 2015. **10**(4): p. 746-761.
28. Singh, A.K., P. Joshi, and G.C. Nandi. *Face recognition with liveness detection using eye and mouth movement*. in *Signal Propagation and Computer Technology (ICSPCT), 2014 International Conference on*. 2014.
29. Sooyeon, K., et al. *Face liveness detection using variable focusing*. in *Biometrics (ICB), 2013 International Conference on*. 2013.
30. Younghwan, K., Y. Jang-Hee, and C. Kyounggho, *A motion and similarity-based fake detection method for biometric face recognition systems*. Consumer Electronics, IEEE Transactions on, 2011. **57**(2): p. 756-762.
31. Tirunagari, S., et al., *Detection of Face Spoofing Using Visual Dynamics*. Information Forensics and Security, IEEE Transactions on, 2015. **10**(4): p. 762-777.
32. Junjie, Y., et al. *Face liveness detection by exploring multiple scenic clues*. in *Control Automation Robotics & Vision (ICARCV), 2012 12th International Conference on*. 2012.
33. Wei, B., et al. *A liveness detection method for face recognition based on optical flow field*. in *Image Analysis and Signal Processing, 2009. IASP 2009. International Conference on*. 2009.
34. Anjos, A., M.M. Chakka, and S. Marcel, *Motion-based counter-measures to photo attacks in face recognition*. Biometrics, IET, 2014. **3**(3): p. 147-158.
35. Kollreider, K., H. Fronthaler, and J. Bigun, *Non-intrusive liveness detection by face images*. Image and Vision Computing, 2009. **27**(3): p. 233-244.
36. Erdogmus, N. and S. Marcel. *Spoofing in 2D face recognition with 3D masks and anti-spoofing with Kinect*. in *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*. 2013.
37. Erdogmus, N. and S. Marcel, *Spoofing Face Recognition With 3D Masks*. Information Forensics and Security, IEEE Transactions on, 2014. **9**(7): p. 1084-1097.
38. Kose, N. and J.L. Dugelay. *Countermeasure for the protection of face recognition systems against mask attacks*. in *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*. 2013.
39. Wallace, R., et al. *Inter-session variability modelling and joint factor analysis for face authentication*. in *Biometrics (IJCB), 2011 International Joint Conference on*. 2011. IEEE.
40. Chen, T., et al., *Total variation models for variable lighting face recognition*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2006. **28**(9): p. 1519-1524.

# Cryptanalysis of Simplified-AES Encrypted Communication

Vimalathithan.R

Dept. of Electronics and Communication  
Engg.  
Karpagam College of Engineering  
Coimbatore, India

M. Omana, C. Metra

Dept. of Electrical, Electronic and Information  
Engineering  
University of Bologna  
Bologna, Italy

D. Rossi

Dept. of Electronics and Computer  
Science University of Southampton  
Southampton, UK

M. L. Valarmathi

Dept. Computer Science  
Government College of Technology  
Coimbatore, India

**Abstract**—Genetic algorithm based Cryptanalysis has gained considerable attention due to its fast convergence time. This paper proposes a Genetic Algorithm (GA) based cryptanalysis scheme for breaking the key employed in Simplified- AES. Our proposed GA allows us to break the key using a Known Plaintext attack requiring a lower number of Plaintext-Ciphertext pairs compared to existing solutions. Moreover, our approach allows us to break the S-AES key using also a Ciphertext-only attack. As far as we are concerned, it is the first time that GAs are used to perform this kind of attack on S-AES. Experimental results prove that our proposed fitness function along with GA have drastically reduced the search space by a factor of 10 in case of Known plain text and 1.8 in case of Ciphertext only attack.

**Index Terms**— Cryptanalysis, Genetic Algorithm, Plaintext, Ciphertext, Simplified-AES.

## I. INTRODUCTION

Cryptography plays a vital role both in wired and in wireless networks. This is especially the case for wireless networks in which, being data transmitted in free space, anyone can access them, thus mandating cryptography to provide security in the communication among nodes [1-3].

Cryptography is the study of methods to obtain messages in disguised forms by using a secret key that is shared between the sender and the receivers. The disguise can be removed, and the message can be retrieved only by intended recipients, who own the secret key. The message to be sent is called plaintext, while the disguised message is called Ciphertext. If Cryptography is the art of making Ciphertext, Cryptanalysis is the art of breaking Ciphertext. Particularly, Cryptanalysis is the study of mathematical techniques that can be employed by an intruder (attacker) to defeat cryptographic algorithms and attack the Ciphertext and retrieve the plaintext, without knowing the secret key [1].

Cryptanalysis is a challenging task. There are several types of attacks that a cryptanalyst may use to break a cipher,

depending upon how much information is available to the attacker. One type of attack is the *Known Plaintext* attack (KPA), in which the attacker has samples of both the plaintext and its corresponding Ciphertext [4].

Another type of attack is the *Ciphertext only* attack (COA), in which the Ciphertext only is available to the cryptanalyst [4-8]. Between the two attacks, the KPA is easier to implement compared to COA, since more information is available to the attacker (both plaintext and Ciphertext pairs) so that the secret key can be more easily retrieved.

Additionally, the computational complexity in attacking the cipher depends not only on the amount of available information, but also on the encryption algorithm. Simplified-Advanced Encryption Standard (S-AES) is a well-known encryption algorithm, frequently used in embedded systems like mobile phones, GPS receivers, etc., which requires low memory and low processor capacity [9]. S-AES is a Non-Feistel Cipher [5] that takes a 16 bit plaintext, 16 bit key and generates a 16 bit Ciphertext. Its encryption uses one pre-round transformation and two round transformations [5].

Several methods have been proposed in the literature to attack S-AES [10-13]. They deal only with KPAs, and this is a strong limitation, since only in very few realistic cases the plaintext and its corresponding cipher text are available.

In 2003, Musa attacked S-AES using Linear and Differential Cryptanalysis [10]. To attack only the (pre-round and) round one in S-AES, 109 plaintext and the corresponding Ciphertext pairs were required. It should be noted that this is a very large number, and it is difficult to obtain in practical applications. Moreover, if the complete S-AES is considered (i.e., also the second round is included), as it is the case in practical applications, the number of plaintext and corresponding Ciphertext pairs required for cryptanalysis increases considerably, making this approach unpractical.

In 2006, Bizaki analyses the complete Mini-AES (S-AES) using linear Cryptanalysis [11]. It has been shown that at least 96 plain text and corresponding cipher text pairs are required

for this type of attack, thus suffering from analogous limitations of [10].

In 2007, Davood attacked Simplified AES with Linear Cryptanalysis using KPA [12]. To break only the first round 116 plaintext and Ciphertext pairs were required, while 548 pairs were required to break also the second round. As previously highlighted, such very large number of pairs is very difficult to be available in practical applications.

In 2009, Simmons proposed a KPA attack to S-AES using Algebraic cryptanalysis [13]. However, in order to apply Algebraic cryptanalysis, a large number of non-linear polynomials have to be constructed, where the variables in the polynomials are unknown key bits, plaintexts and Ciphertext. It is well known that solving a set of non-linear equations is a complex and time consuming task.

Recently, it has been proven that Genetic Algorithms (GAs) can be effectively adopted to retrieve the key used for encrypting messages without searching the entire key space [14]. As known, GAs provide efficient and effective searches in complex space [15,23], they are computationally efficient and can be easily implemented. Starting from an initial random population, GAs efficiently exploit historical information contained in the population. By applying GA parameters on previous population, we obtain a new search space, from which the expected result can be obtained at a faster rate. These GA properties have been exploited in [14] to attack S-DES, by effectively tuning the GA parameters. In this regard, however, it is worth noticing that S-DES can be easily attacked, since the encryption algorithm uses only 10 bit key and does not have any nonlinearity, whereas in case of S-AES, the key size is 16 bit and the algorithm is nonlinear. Therefore, S-AES is more complex and difficult to attack compared to S-DES.

As Clarified above, only KPA (and not COA) has been proposed so far to attack S-AES, since COA is much more complex than KPA. This makes the attack of COA using linear cryptanalysis unfeasible. As an alternative, COA can be carried out by trivial brute-force attack, where the cryptanalysers tries every possible combination of keys until the correct one is identified. However, this type of attack is very time consuming, if lengthy keys are used for encryption. It can become feasible only by using a network of computers and combining their computational strengths, although its cost would be extremely high [4, 5].

In this paper we address the issue of attacking S-AES. We propose a new GA based approach that is able to attack efficiently S-AES, using either KPA or COA, thus overcoming the above mentioned limitations of alternative approaches. As far as we are concerned, this is the first time that GAs are used to perform Cryptanalysis of S-AES.

In case of KPAs, we will show that our approach requires a smaller number of plaintext and Ciphertext pairs, when compared to alternative linear cryptanalysis attacks in [10-13], thus being more suitable to be employed in practical applications. As for COAs, as discussed above, no alternative solution, other than the trivial brute-force attack, does exist. Compared to brute-force attack, we will show that our approach is significantly faster.

The rest of the paper is organized as follows. In Section 2, we recall the basic principles of S-AES and GAs. In Section 3, we describe our proposed GA based approach. In Section 4, we report some experimental results, while Section 5 concludes our paper.

## II. PRELIMINARIES

### A. Basics of Simplified AES

In this section we recall the basics of the S-AES algorithm. More details about S-AES encryption, key expansion and its decryption algorithm can be found in [5, 10].

#### 1) Encryption

S-AES is a Non-Feistel Cipher [5] that takes a 16 bit plaintext, a 16 bit key and generates a 16 bit Ciphertext. S-AES encryption procedure consists of three phases, namely one Pre-round transformation and two Round transformations (referred to as Round 1 and Round 2).

The encryption, key generation and decryption steps are illustrated in Figure 1. The Pre-Round phase uses a single transformation, referred to as Add Round Key. Instead, Round 1 uses the following four transformations: Substitution, Shift Row, Mix Columns and Add Round Key, while Round 2 uses the same transformations as Round 1, with the exception of the Mix Column one.

The 16-bit input plaintext, called *state*, is divided into two-by-two matrix of nibbles, where one nibble is a group of 4 bits. The initial value of the state matrix is the 16-bit plaintext; the state matrix is modified by each subsequent function in the encryption process, producing the 16-bit Ciphertext after the last function.

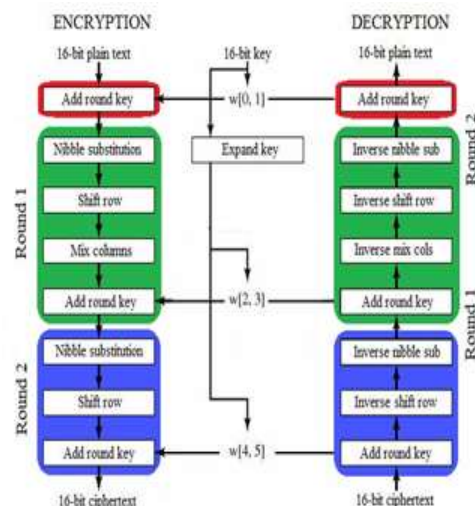


Figure 1. Encryption, Key Generation and Decryption Algorithm for Simplified-AES.

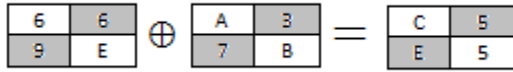


Figure 2. Example of Add Round Key transformation.

Each round takes a state and creates another state to be used for the next round by applying the respective transformations. In the Pre-Round, the Add Round Key transformation is applied. It consists of the bitwise XOR of the 16-bit state matrix and the 16-bit round key. As shown in Figure 2, it can also be viewed as a nibble-wise (or bit-wise) matrix addition over the GF (2<sup>4</sup>) field.

All the transformation used in Round 1 and Round 2 can be described as follows.

a) *Substitution*- As a first step in Round 1, Substitution is performed for each nibble. The nibble substitution function is based on a simple lookup table, denoted as substitution table, or S-box. An S-box is a 4 x 4 matrix of nibble values that contains a permutation of all possible 4-bit values. Each individual nibble of the state matrix is mapped into a new nibble in the following way: The leftmost 2 bits of the nibble are used as a row index and the rightmost 2 bits are used as a column index. These row and column indexes identify into the S-box a unique 4-bit output value (the new nibble). The transformation definitely provides confusion effect, which make the relationship between the statistics of the ciphertext and the key as complex as possible, again to baffle attempts to discover the key [5]. As an example, the S-box used for the encryption is:

$$S = \begin{bmatrix} 9 & 4 & A & B \\ D & 1 & 8 & 5 \\ 6 & 2 & 0 & 3 \\ C & E & F & 7 \end{bmatrix}$$

b) *Shift Row*- The shift row function performs a one-nibble circular shift of the second row of the state matrix, while the first row is not altered.

c) *Mix Columns*- As a third step, Mix Column is carried out. It changes the content of each nibble, by taking 2 nibbles at a time and combining them to create 2 new nibbles. To guarantee that each new nibble is different, even though the old nibbles were the same, the combination process first multiplies each nibble by a different constant, then it mixes them. The mixing can be performed by matrix multiplications. Multiplication of bytes is done in GF (2<sup>4</sup>), with modules (x<sup>4</sup>+x+1) or (10011).

d) *AddRound Key*- Finally, Add Round Key is performed. Analogously to the operation performed during the Pre-Round, it involves the cipher key.

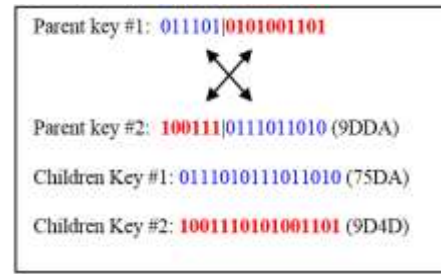


Figure 3. Example of Crossover transformation.

## 2) Decryption

Decryption is encryption reverse process. It takes a 16 bit ciphertext, the 16 bit key, and generates the original 16 bit plaintext. Similarly to encryption, decryption uses one pre-round and two round transformations, as shown in Figure 1. The processes performed during decryption are the inverse of those employed in encryption [5].

## 3) Key Generation

To increase the security of S-AES, starting from the original 16 bit cipher key, three additional round keys are generated, by applying a proper key generation algorithm [5]. This allows to use a different key for each round. The same keys used for encryption are used also for decryption. Of course, in this latter case, the order of the used keys is reversed.

## B. Genetic Algorithms

Genetic algorithms are inspired by Darwin's theory of evolution [17-21]. GAs provides effective and efficient searches in complex spaces. They are computationally efficient, and are not limited by any restrictions on the search space like random search methods that, instead, work properly only within certain boundaries, or under specific limiting conditions. Moreover, with random search methods, the algorithm may get stuck into the problem of local minima, thereby increasing computational time [15].

The terms used in GA are:

- Gene – A single bit in the chromosome
- Chromosome (Individual) - Any Possible Solution
- Population - Group of Chromosomes
- Search Space - All possible solutions to the problem
- Fitness Value - A function to evaluate performance
- Generations – Number of Iterations.

GAs are preferable to random searches when the search space is large, complex or unknown, and when mathematical analysis is unavailable.

Now let us briefly explain how GAs work. As a first step, a population representing chromosomes is randomly created. Then, the individuals in the population are evaluated using a proper fitness function, and a value is assigned to each individual, based on how efficiently it performs the task. Two individuals are selected based on their fitness value, and the one with the higher fitness wins. The individuals which win are employed to reproduce and create one or more offsprings, after

which the offsprings are randomly mutated. This process continues until a suitable solution is found, or certain number of generations has passed.

A simple GA that yields good results in many practical problems is composed by three operators: *Selection* (*Reproduction*), *Crossover* and *Mutation*.

Selection strategies determine which chromosome will take part in the evolution process. The different Selection strategies are *Tournament Selection*, *Population Decimation* and *Proportionate Selection* [15]. In Tournament Selection, two individuals are randomly selected and the one with the highest fitness wins. This process continues until the required number of chromosomes is obtained. Details about other selection strategies can be found in [15].

After *Selection*, *Mating* is performed. While Selection addresses the issue of selecting which individuals will take part in the evolution process, *Mating* selects which two parent chromosomes will mate with one another. Several Mating schemes are possible. They include Best-Mate-Worst (BMW), Adjacent Fitness Pairing (AFP) and Emperor Selective Mating (ESM) scheme. In BMW mating scheme, as the name indicates, the chromosome with the highest fitness mates with the chromosome with the lowest fitness. In case of ASP, the two keys with the lowest fitness mate together, the keys with the next two lowest fitnesses mate together, etc. In ESM, the highest ranked individual mates with the second highest, fourth highest, etc. individuals (that is, with all even order individuals), while the third, fifth, etc. highest individuals (that is, those with odd order) remain unchanged.

The next operation performed by GAs is crossover, which selects genes from parent chromosomes and creates a new offspring. Crossover is followed by mutation, which randomly changes one or more bits in the chromosome.

### III. PROPOSED APPROACH FOR ATTACKING S-AES USING GENETIC ALGORITHMS

In this section we propose the use of GA in Cryptanalysis, in order to break the Cipher key for KPA and COA.

#### A. Adoption of GA in Cryptanalysis

As introduced in Section 2.2, GA starts with a set of solutions constituted by chromosomes, called *initial population*. In our case, a chromosome represents a key, and the length of the key corresponds to the size of the chromosome. A chromosome size of 16 bits is considered, since this is the size of a S-AES key. From the old set of used keys, a new set of keys is generated, in order to form a new solution. The new set of keys may be a better solution (closer to the actual key) than the older one.

Given a set of keys, each one with a fitness value, GA generates a new set of keys using GA parameters. Reproduction or Selection strategies determine which key will take part in the evolution process to make up the next generation, in terms of mating with other keys. Among the three different selection strategies, as discussed in Section 2.2, *Tournament Selection* is the best suited for cryptanalysis [14].

After the selection process is accomplished, the mating operation is performed. Among the three possible mating schemes, the Best-Mate Worst scheme is the preferred one in cryptanalysis, mainly due to avalanche effect in block ciphers [4], where a small change in the plaintext, or key, creates a significant change in the ciphertext.

After the key is selected by the BMW mating scheme, the *crossover* operator is applied. Consider the following two parent keys:

- Parent key #1 : 0111010101001101 (754D)
- Parent key #2 : 1001110111011010 (9DDA)

Crossover, mates the two parent keys to produce two offsprings (children keys). To perform crossover, the crossover point, that is the point at which the key will be split, has to be selected.

We consider that case of random crossover, since, as shown later, it performs better than the other crossover types in case of cryptanalysis. The crossover point  $k$  is chosen randomly in the range  $[0, keylength]$ . If  $k$  is equal to  $keylength$  (or 0), then no crossover will occur. For example, a crossover point of  $0.5*keylength$  would indicate that the parent keys would be cut in half. In case of Uniform Crossover, the value  $k$  is fixed. The example in Figure 3 shows the case where the crossover point  $k$  is 6, for 16 bit parent keys.

The two newly generated keys may have a better fitness value than their parent keys: in this case the evolution process continues. Instead, if the children keys have a worst fitness value than their parent keys, half of the population of parent keys, and half of the population of the children keys are selected as new parent keys for the next generation, and the evaluation process continues.

It should be noted that we have selected a single crossover point, rather than two crossover points, since it has been verified that it produces better results [22].

Finally, the operation of mutation is performed. The mutation operator randomly changes one or more bits in a key, thus preventing the population from missing the optimal fitness value [15]. In the example below, the tenth bit equal to '1' is mutated to a '0' to obtain a new key.

- Before mutation : 1001110111011010 (9DDA)
- After mutation : 1001110110011010 (9D9A)

#### B. Fitness Function

As introduced in Section 2.2, to evaluate the performance of GA, a proper fitness function has to be defined. Two different fitness functions for the two different types of attacks have been defined. As described in details in the following subsections, in case of KPA, the chosen fitness function is the correlation between the known ciphertext and the generated ciphertext, whereas in the case of COA, a more complex fitness function has been developed, which employs letters' frequency.

1) *Known Plaintext attack*- In KPA, the attacker takes advantage of having samples of the plaintext and its corresponding ciphertext. For example, encrypted file archives, such as ZIP [24], as well as encrypted system files on hard-disk [24], are prone to this kinds of attacks.



An attacker with an encrypted zip file needs only one unencrypted file (known plaintext) from the archive, which can be guessed from the knowledge of the file name [24]. Then, the key required to decrypt the entire archive can be instantly found. As another example, we can consider the case of encrypted e-mails, where the e-mail headers are sent in both unencrypted (plaintext) and encrypted form (Ciphertext), thus making KPA possible [25].

The fitness function for KPA must relate the plaintext, the ciphertext, and the key used for encryption. Hence, the correlation function between the known ciphertext and the generated ciphertext can be used as fitness function. It is given by:

$$F_{kp} = \#(C_k \oplus C_g) / 16, \quad (1)$$

where  $C_k$  is the known ciphertext (generated using the actual key),  $C_g$  is the generated ciphertext (generated using the trial key),  $\oplus$  represents the bit-wise Xor operation,  $\#$  denotes the number of ones in  $C_k \oplus C_g$ . In (1) the normalization factor is 16, since, as introduced above, we are considering the case of 16 data bits.

The range of  $F_{kp}$  is (0, 1). Particularly,  $F_{kp}$  is 0, if all bits in  $C_k$  and  $C_g$  are identical, while  $F_{kp}$  is 1, if all bits in  $C_k$  and  $C_g$  are different. The actual key is found when  $C_k$  and  $C_g$  are identical, so our goal is to minimize the fitness function ( $F_{kpmin}=0$ ).

In order to compute the fitness value, the known plaintext is encrypted using randomly generated keys to generate the ciphertext  $C_g$ . Once  $C_g$  is generated, a bit-wise XOR operation between  $C_k$  and  $C_g$  is performed, and the fitness function is evaluated.

2) *Ciphertext-Only attack*- COA is the attack where the attacker does not know anything about the content of the message, and only has a sample of ciphertext. The success of such an attack increases with the number of available samples of ciphertext, provided that each sample has been encrypted using the same algorithm and key. Therefore, COA is one of the most difficult attacks to be performed.

Since in COA the ciphertext is the only available information, the fitness function defined in the previous subsection cannot be used, and a new fitness function has to be defined.

As an example, here we assume that the plaintexts were constructed using the English language. In order to derive the fitness function, the ciphertext is first decrypted, then letter frequency analysis is performed from the obtained plaintext. The fitness function for COA can be given by equation (2):

$$F_{cip} = \alpha \sum_{i \in \tilde{A}} |K(i)^u - D(i)^u| + \beta \sum_{(i,j) \in \tilde{A}} |K(i,j)^b - D(i,j)^b| + \gamma \sum_{(i,j,k) \in \tilde{A}} |K(i,j,k)^t - D(i,j,k)^t| \quad (2)$$

Where  $\tilde{A}$  denotes the language alphabet {A,B...Z, \_} for English (where \_ represents the space symbol); K and D are the known language statistics and decrypted message statistics, respectively;  $u$ ,  $b$ , and  $t$  denote the unigram, digram (two letter

TABLE I  
CONSIDERED DIGRAMS AND TRIGRAMS

Considered Digrams	Considered Trigrams
TH, HE, IN, ER, AN, RE, ED, ON, ES, AT, TO, NT,ND, HA, EA, OU, IS, IT, TI, ET, AR, TE, SE, HI, OF, AS, OR.	THE, ING, HER, ERE, AND, THA, WAS, FOR, ION, HAS, MEN, NCE

combinations) and trigram (three letter combinations) statistics (i.e., their occurrence frequencies), respectively. For example, the known statistics for the frequency of occurrence of the letter 'E' in an English text is 12.7%, while it is of the 9.1% for the letter 'T'; the digram frequency for 'TH' is 3.21%, while it is of the 3.05% for 'HE'. The frequency statistics for all other digrams and trigrams can be found in [6]. The unigram statistics  $u$  can be computed by counting the number of occurrences of each character, and dividing it by the total number of characters. An analogous approach can be followed to compute the statistics of  $b$  and  $t$ . Among all possible  $27$  unigrams,  $27^2$  digrams and  $27^3$  trigrams, we considered all unigrams, and a few digrams and trigrams, in order to limit our problem computational complexity. Particularly, we have considered only the most frequently occurring digrams and trigrams, which are reported in Table I.

Finally  $\alpha$ ,  $\beta$  and  $\gamma$  are the weights assigning different priorities to each of the three statistics. The normalization condition  $\alpha+\beta+\gamma=1$  holds true. In the considered case, the parameter  $\alpha$ ,  $\beta$  and  $\gamma$  takes the values 0.2, 0.4 and 0.4, respectively, as reported in [16].

As discussed in Section 2.2, the fitness function has to be evaluated in order to find the key. First, a randomly generated key is used to decrypt the cipher text, i.e., to obtain the plaintext. Once the plaintext is obtained, the fitness function is evaluated by computing unigram, digram and trigram statistics (i.e., their occurrence frequencies). These decrypted message statistics are subtracted from the known language statistics. The absolute values of each frequency statistics are scaled by the weighting coefficients, and summed to each other to compute the fitness value, as shown in (2).

Similarly to the KPA case, in COA the goal is to minimize the fitness function. To set the minimum fitness value, standard size (length) text files were extracted from English novels and

Cryptography books (texts only), to then compute the average of the cost for all considered text files. Even without knowing the message type contained in the ciphertext file (i.e., whether the content of the file is Standard English, or technical content) the computed average fitness value can be used as minimum fitness value. The range of minimum fitness value which has been found with this analysis is 0.1-0.2.

### 3) Algorithm for finding the key using GA

In order to find the key in KPA and COA, the following steps have to be performed. The entire process is shown in Figure 4, and described hereinafter.

1. Generate randomly initial keys. The number of keys considered initially represents the population size. The results show that it is better to consider a low population size, and increase the number of generations. This way, a higher number of crossovers takes place, thereby increasing the crossover rate and, as a final result, reducing the key search space.
2. Using the randomly generated keys:
  - a. In case of KPA, encrypt the known plaintext to generate the ciphertext, to then compute the Fitness function  $F_{kp}$ .
  - b. In case of COA, decrypt the known cipher text and evaluate the Fitness function  $F_{cip}$  from the plain text, by computing the letter frequencies.
3. Compare the computed fitness value with the expected minimal fitness value. If it is lower than, or equal to the minimum fitness value, we can conclude that the corresponding key with minimum fitness is the optimal key. An additional step that is applicable to KPA only consists in checking the correctness of the key by comparing other pairs of known plaintext and ciphertext. This step is represented by the dotted block (check for key confirmation) in Figure 2. If the condition in step 3 is satisfied, stop the algorithm, else continue to step 4.
4. If the computed fitness is not less than or equal to minimum fitness, then apply GA parameters and continue the evolution process.
5. Select the parent keys to generate a new set of children keys, using the selection strategies defined in Section 3. Do the crossover (random crossover is preferable)
6. Perform the Mutation.
7. For the newly generated keys, compute the fitness function and go to step 3.
8. Repeat the step 2 to 7, until the minimum fitness value is achieved, or the chosen maximum number of generations is reached.

If the maximum number of generations is reached, then then key with the minimum fitness value in the final generation is considered as the optimal key. All the described processes are shown in Figure 4.

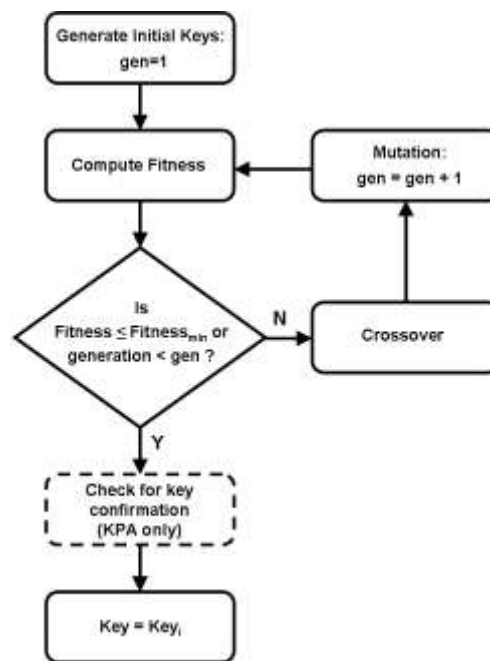


Figure 4. GA cycle for Cryptanalysis.

## IV. EXPERIMENTAL EVALUATION AND COMPARISON

The proposed algorithm has been implemented using Matlab on an Intel PIV processor. The performances of the proposed GA-based approach in attacking the cipher key have been analyzed.

For KPA, the selected GA parameters are shown below.

- Crossover Type: Random
- Mating scheme: Best-Mate-Worst
- Mutation rate: 0.015

The total number of generations depends on the initial population size. The initial population and generation is taken in such a way that the total key search space is set to 16,000, in order to keep the search space smaller compared to the brute-force search space, at least by a factor of 4. For instance, if the initial population size is taken as  $2^2$ , the number of generation is 4000. Similarly, when the initial population size is  $2^5$ , the number of generations is 500.

The known plaintext and ciphertext pairs used in our experiments are shown in Table II. For instance, the known plaintext 0110 1111 0110 1011 (6F6B), which represents the ASCII bit pattern for the text 'ok', and its corresponding known ciphertext 0000 0111 0011 1011 (0738) are considered. Using this pair, the optimum key is obtained by using randomly generated keys and by applying our proposed GA based approach.

Then the obtained key is used to encrypt other known plaintexts, in order to check the correctness of the found key. For instance, assume that the key used for encryption is 1010 0111 0011 1011 (A73B), but the key obtained is 1010 0100 0101 1111 (A45F), as shown in Table II. This obtained key is used to encrypt the other plain text 0110 1000 0110 1001 (6869)



and 0110 1001 0111 0011 (6973), which results in a ciphertext that differs from its corresponding correct ciphertext pair 1100 1011 1001 1010 (CB9A) and 1111 0100 1101 0110 (F4D6). This confirms that A45F is not the actual key. Afterwards, by using the other set of known plaintext and ciphertext pairs, and by applying GA, the new key 1010 0111 0011 1011 (A73B) is found and checked for correctness. The resulted new key, when used to encrypt another known plaintext, results in a ciphertext which is identical to the corresponding known ciphertext, thus confirming that the key 1010 0111 0011 1011 (A73B) is the actual key. From Table II, by using the known plaintext and ciphertext pairs and by analyzing the obtained key in setup 1, the optimal key turns out to be 1010 0111 0011 1011 (A73B). This procedure is carried out for another two sets of three plaintext and ciphertext pairs, as is shown in Table II.

TABLE II : EXPERIMENTAL RESULTS FOR KNOWN PLAINTEXT ATTACK TO FIND OPTIMAL KEY.

Case	Key used	Known Plaintext	Known Ciphertext	Key Found
1	A73B	6F6B	0738	A45F
		6869	CB9A	A73B
		6973	F4D6	A73B
2	A73B	616E	5547	A73B
		6974	D4DE	A73B
		6966	EA54	A73B
3	A73B	6279	C36A	A73B
		6F6E	5737	A73B
		696E	5A57	A73B

Table III shows that, if the initial population size is small, then the key search space is also small, and the key can be found fast, this is due to the fact that crossover rate is high i.e., in each generation the new chromosomes were created by crossover thereby searching with new keys and make the algorithm to converge quickly. Also if the population size is too low then the algorithm converges slowly. Figure 5 shows how the fitness value converges depending on the number of generations, considering the first case shown in Table III as an example. Particularly, the fitness value converges to zero after 1726 generations, showing that the key search space size is equal to 6905 only, thus being considerably lower (by a factor of 10) than the brute-force search space size (which is equal to  $2^{16}$ ).

TABLE III  
EXPERIMENTAL RESULTS FOR KNOWN PLAINTEXT ONLY ATTACK

Key used	Pop. size	Key searched	Key found	Success Key bits
A73B	2 <sup>2</sup>	6905	0738	16
A73B	2 <sup>5</sup>	11136	CB9A	16
A73B	2 <sup>1</sup>	17232	A73B	16
4E26	2 <sup>2</sup>	4220	4E26	16

Our proposed approach allows us to find the key by using three plaintext-ciphertext pairs only, whereas in other cases the number of plaintext-ciphertext pairs required were reported in Table IV, as highlighted in Sect. 1, they were a very large and difficult to be obtained number.

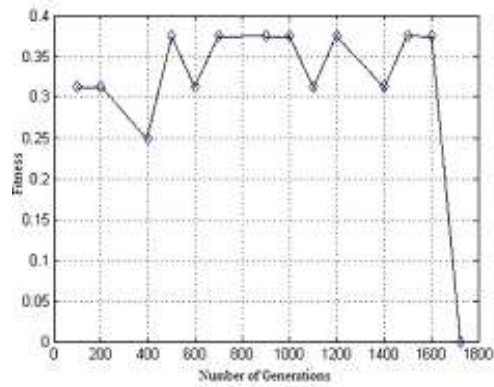


Figure 5. Fitness as a function of Number of Generations for KPA.

TABLE IV  
NUMBER OF PLAINTEXT-CYPHERTEXT REQUIRED FOR ATTACKING

Technology	Rounds attacked	Number of plaintext-ciphertext pair required
Linear Cryptanalysis Musa [10]	Round1	109
Linear Cryptanalysis Davood [12]	Round1	116
	Round1 & Round2	548
Linear Cryptanalysis Bizaki [11]	Round1 & Round2	96
Our approach using GA	Round1 & Round2	3

TABLE V  
EXPERIMENTAL RESULTS FOR COA USING VARIOUS KEYS

Key used	Selection type	Crossover type	Mutation rate	Key found	Key searched
A73B	Best-mate worst	Random	0.015	A73B	18,281
A73B	Best-mate worst	Uniform	0.015	A73B	20,349
A73B	Best-mate worst	Random	0.0	A73B	19,125
B6E7	Best-mate worst	Uniform	0.0	B6E7	22,148
A420	Adjacent Pair	Random	0.0	A420	21,092

For COA, the initial population size is set to 32, with a chromosome size of 16 bits, that is, 32 sets of 16 bit keys that are randomly taken. The total number of generations is taken as 1000, i.e., the key search space size is set to a maximum of 32,000, in order to keep the COA search space lower enough (approximately one half) than the brute-force search space size. As discussed in the previous section, the known cipher text is decrypted using the initial keys, and the fitness function is calculated using equation (2) for each key. For each generation, a final solution is produced, and the optimum solution (i.e., the key) is found, based on the minimum fitness value.

In order to evaluate different trade-offs, the GA parameters have been set to different values, and the results have been compared. Table V shows the results for cryptanalysis using GA, where the size of the considered ciphertext file is equal to 1000 characters. The results highlight that all 16 bits of the key are effectively found. Particularly, by considering the first two lines, it can be seen that, if the random crossover point is used,

TABLE VI  
EXPERIMENTAL RESULTS FOR VARIOUS WITH BEST GA PARAMETERS

GA Parameters	Key used & found	Key searched
Initial Population-32	A73B	18,281
Chromosome size-16	A73B	18,529
No. of iteration -1000	A73B	18,764
Crossover – BMW	A73B	19,020
Mutation rate -0.015	A73B	19,020

the algorithm convergences faster than in the uniform crossover. In fact, as reported in the last column of Table V, the number of searched keys is equal to 20349, in case of uniform crossover, while this number is reduced to 18281, when random crossover is used, with a 10% reduction in the number of searched keys.

Additionally, by considering the next two cases in Table V, if the mutation rate is set to zero, the key can be recovered with a slight increase in the search space. The case of BMW as selection type, with random crossover and mutation rate equal to 0.015, is the best case in terms of number of searched keys. In this case, the key is retrieved in 600 generations, with a key search space slightly larger than approximately 18000. It should be noted that, in the brute-force attack, the search space is  $2^{16}$ , in the worst case. Thus GA reduces the search space by a factor of approximately 3.6, which represents a very high improvement in cryptanalysis. Table VI show the results for attacking the key by considering best GA parameters as specified in the first row of Table V for various initial populations. On average, by using GA, the key search space is reduced by a factor of 1.8 when compared to the average case of brute-force attack where the search space is  $2^{15}$ .

Figure 6, shows how the fitness value converges as a function of the number of generations, by considering the GA parameters reported in the first row of Table V. It shows also how the fitness value depends on the amount of cipher text, considering three different examples for the ciphertext size (100, 500 and 1000 characters). As can be seen, if the size of the cipher text increases, then the algorithm converges quickly with a lower number of generations, while the number of generations required to reach the desired fitness value increases in case of lower ciphertext size. Particularly, in the case of 1000 Ciphertext, the algorithm converges to the desired fitness value after 600 generations. Instead, for the case of 100 ciphertexts only, the algorithm takes 820 generations to converge to the final value, which is higher than the optimal value.

The convergence of the algorithm as a function of the size of the ciphertext is more clearly shown in Figure 7. If the size of the ciphertext is small, the respective decrypted plain text contains little information about the letter frequency and it is difficult to compare with the standard letter frequency analysis, since the latter is usually constructed using large file size.

Hence the algorithm needs more generations to converge. Reversely, if the size of the ciphertext is large, the fitness function can be computed more easily, since more letter frequency information can be extracted from the decrypted plain text. As a result, the key is recovered quickly with fewer generations. In this regards, it should be noted that if the size of

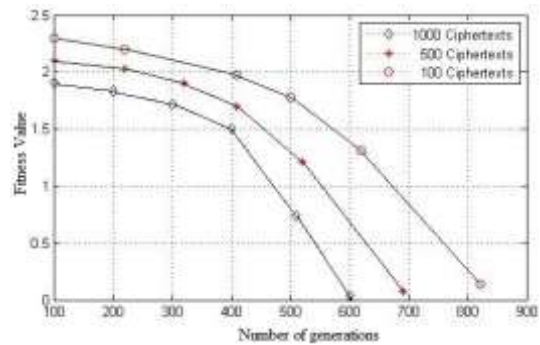


Figure 6. Fitness as a function of number of Generations.

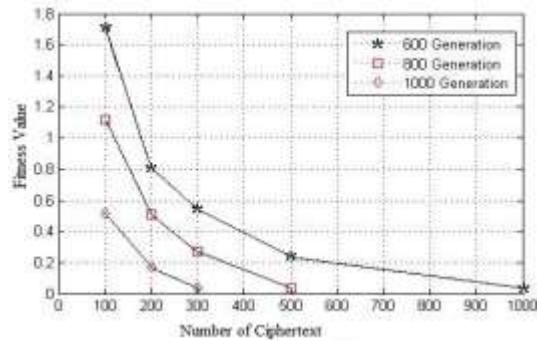


Figure 7. Fitness as a function of number of cipher text.

the ciphertext increases, then the computational time for one decryption increases as well, but the number of generations decreases, and also the algorithm converges with a reduced key search space. On the whole, the convergence of the algorithm is faster. In all cases, the key was successfully found, requiring a minimum of 100 ciphertexts to compute effectively the letter frequency analysis.

## V. CONCLUSION

A new GA based approach for attacking Simplified-AES by KPA and COA has been proposed. Our experimental results show that the proposed algorithm can effectively break the key. In case of KPA, three pairs of plaintext and ciphertext suffice to break the key, whereas in case of alternative Linear cryptanalysis, 512 plaintext cipher text pairs are required, which is a very large and very difficult to be obtained number. Differently from previous solutions, our proposed algorithm allows to break the key successfully also by COA. In this case, our algorithm allows to reduce dramatically the key search space, compared to the existing (and unique) alternate brute-force attack. The results says that the proposed research have drastically reduced the search space by a factor of 10 and 1.8 in case of Known plain text and ciphertext only attack respectively. Though Simplified-AES is simpler than AES, our proposed approach paves the way to attack AES. In fact, the fitness function used for KPAs can be directly applied to other block ciphers, like AES. Instead, the fitness function used for COAs in Simplified-AES is not appropriate for AES, as the compilation of frequency statistics becomes infeasible when the

number of bits is increased to 128 bits. However, the approach followed to develop the fitness function for COAs in Simplified-AES give cryptanalysers useful insight to attack AES effectively using COA.

#### REFERENCES

- [1] N. Koblitz, "A course in Number Theory and Cryptography" Springer International Edition, 2008.
- [2] D.Rossi, M.Omaña, D.Giaffreda, C.Metra, "Secure Communication Protocol for wireless Networks". *IEEE East- West Design and Test Symposium 2010*.
- [3] <http://teal.gmu.edu/crypto/rijndael.htm>
- [4] W. Stallings, "Cryptography and Network Security Principles and Practices" Pearson Education, 2004.
- [5] B. A. Forouzan, 'Cryptography and Network Security" Tata Mc Graw hill Education, 2nd edition, 2008.
- [6] A.Menezes, P.Vanoorschot, S.Vanstone, 'Hand Book of Applied Cryptography' CRC Press, 1996.
- [7] H.F.Gaines, Cryptanalysis, "A study of Ciphers and their solution", Dover Publications, NewYork.
- [8] M.Stamp, R.M. Low, "Applied Cryptanalysis", Wiley-Interscience, A John wiley & sons, Inc Publications.
- [9] S.J. Manangi, P. Chaurasia, M.P. Singh, "Simplified AES for Low Memory Embedded Processors", *Global Journal of Computer of Computer Science and Technology*, Vol.10 Issue 14, Nov 2010, pp. 7-11.
- [10] M. A. Musa, E.F. Schaefer, S. Wedig, "A Simplified AES algorithm and its linear and Differential Cryptanalysis", *Cryptologia*; April 2003, pp. 148-177.
- [11] H.K. Bizaki, S. David Mansoor, A. Falahati, "Linear Cryptanalysis on Second Round Mini-AES", *International Conference on Information and Communication Technologies*, 2006, pp. 1958-1962
- [12] S.D. Mansoori, H.Khaleghei Bizaki, "On the Vulnerability of Simplified AES Algorithm Against Linear Cryptanalysis", *International Journal of Computer Science and Network Security*, Vol.7, No.7, July 2007, pp. 257-263.
- [13] S.Simmons, "Algebraic cryptanalysis of Simplified AES", *Cryptologia*, Oct 2009, pp. 305-314.
- [14] Vimalathithan .R, M.L. Valarmathi, "Cryptanalysis of S-DES Using Genetic Algorithm", *International Journal of Recent Trends in Engineering*, Vol2, No.4, November 2009, pp. 76-79.
- [15] D.E. Goldberg, "Genetic Algorithm in Search, Optimization and Machine Learning", Boston, Addison-Wesley, 1999.
- [16] Nalini, "Cryptanalysis of Simplified data encryption standard via Optimization heuristics", *International Journal of Computer science Network and Security*, vol 6, No 1B, Jan 2006.
- [17] Davis.L. "Handbook of Genetic Algorithm", Van Nostrand Reinhold, New York.
- [18] R. Collin, J. Reeves, E. Rowe, "Genetic Algorithms-Principles and Perspectives,A guide to GA theory", Kluwer Academic Publishers.
- [19] M.Mitchell, "An Introduction to Genetic Algorithms", First MIT press paperback edition, 1998.
- [20] N. Nedjah, A. Abraham, L. de Macedo Mourelle, "Genetic Systems Programming", *Studies in Computational Intelligence*, Vol 13, 2006.
- [21] N Nedjah, A Abraham, L de Macedo Mourelle, "Computational Intelligence in Information Assurance and Security", *Studies in Computational Intelligence*, Vol 57, 2007.
- [22] R.L. Haupt, S.E. Haupt, "Practical Genetic Algorithms", 2nd edition, Wiley, 2004.
- [23] Poonam Garg, "Evolutionary Computation Algorithms for Cryptanalysis: A Study" *International Journal of Computer Science and Information Security*, Vol. 7, No. 1, 2010
- [24] [http://en.wikipedia.org/wiki/Known-plaintext\\_attack](http://en.wikipedia.org/wiki/Known-plaintext_attack).
- [25] <http://technet.microsoft.com/en-us/library/cc962032.aspx>

## Risk Assessment in Hajj Event - Based on Information Leakage

Asif Bhat

Department of Information Technology  
International Islamic University Malaysia  
Kuala Lumpur Malaysia

Haimi Ardiansyah

Department of Information Technology  
International Islamic University Malaysia  
Kuala Lumpur Malaysia

Said KH. Ally

Department of Information Technology  
International Islamic University Malaysia  
Kuala Lumpur Malaysia

Jamaluddin Ibrahim

Department of Information Technology  
International Islamic University Malaysia  
Kuala Lumpur Malaysia

**Abstract**---Annually, millions of Muslims embark on a religious pilgrimage called the "Hajj" to Mecca in Saudi Arabia. Management of Hajj activities is a very complex task for Saudi Arabian authorities and Hajj organizers due to the large number of pilgrims, short period of Hajj and the specific geographical area for the movement of pilgrims. The mass migration during the Hajj is unparalleled in scale, and pilgrims face numerous problems. Including RFID tags there are many types of identification and sensor devices developed for efficient use. Such technologies can be used together with the database systems and can be extremely useful in improving the Hajj management. The information provided by the pilgrims can be organised in the Hajj database and can be used to effectively identify individuals. The current system of data management is mostly manual, leading to various leaks. As more of the sensitive data gets exposed to a variety of health care providers, merchants, social sites, employers and so on, there is a higher chance of Risk. An adversary can "connect the dots" and piece together the information, leading to even more loss of privacy. Risk assessment is currently used as a key technique for managing Information Security. Every organization is implementing the risk management methods. Risk assessment is a part of this superset, Risk Management. While security risk assessment is an important step in the security risk management process, this paper will focus only on the Risk assessment.

**Keywords:** Hajj, Information Leakage, Risk Assessment.

### I. INTRODUCTION

The Hajj (Arabic: حج Haġġ "Pilgrimage") is an Islamic pilgrimage to Mecca and the largest gathering of

Muslim people in the world every year. It is one of the five pillars of Islam, and a religious duty which must be carried out by every able-bodied Muslim who can afford to do so at least once in his or her lifetime. Hajj is a unique gathering of its kind and poses a challenge to its organisers. Management of the annual pilgrimage to Mecca known as Hajj is a very complex task. Recently many types of identification and sensor devices, including RFID tags [1], have been developed. Such technologies, together with the use of database can be extremely useful in improving the Hajj management. Information leakage is a real and growing problem. Every month, news about another organization leaking confidential information becomes public. These are the known cases that have a visible impact.

Many similar incidents occur daily and the vast majority of information leaks are accidental: it is not solely the result of intentional, harmful actions. Unintentional data loss is perhaps more dangerous because those affected are not necessarily aware of, or able to act on, the problem. Aside from any other impact, information loss may represent a very high cost for organizations. Information loss has both direct and indirect costs: the intellectual property or industrial information itself together with the cost of handling the consequences of its loss. Indirect costs include: loss of credibility, erosion of competitive advantage and regulatory transgressions [2].

The growing awareness of the risks of information leakage was sparked by a series of corporate scandals in which confidential information was disclosed. As the majority of those cases demonstrate, such breaches are often not the result of malicious wrongdoing, but

rather employees who unknowingly put their companies at risk. This may occur as employees send out email messages that contain files or content that they are not aware is confidential. Another example is employees delivering confidential files to their web-based email boxes, or copying files to mobile devices, and thus exposing them to untrusted environments. This may lead them to Risk assessment. Risk assessment is the process where you [3]:

- Identify hazards.
- Analyse or evaluate the risk associated with that hazard.
- Determine appropriate ways to eliminate or control the hazard.

In practical terms, a risk assessment is a thorough look at your workplace to identify those things, situations, processes, etc. that may cause harm, particularly to people. After identification is made, you evaluate how likely and severe the risk is, and then decides what measures should be in place to effectively prevent or control the harm from happening [4]. The aim of the risk assessment process is to remove a hazard or reduce the level of its risk by adding precautions or control measures, as necessary. By doing so, you have created a safer and healthier workplace [5].

## II. INFORMATION LEAKAGE IN HAJJ EVENT

Hajj is performed every year by pilgrims from 150 countries, but the database that is available in respect of these pilgrims is still in its simplest forms. Hence, there is a dire need for expanding and extending the available data, which would enable the Ministry to improve its performance in dealing with the special circumstances of the pilgrims of each country and to raise the standard of its service plans.

The Ministry is seeking to realize that through the establishment of an information centre as part of its set up and to design programs for benefiting from information programs and banks connected with pilgrims in the countries of origin and in international organizations [6]. This also applies in respect to Umrah performers and visitors to the Prophet's (saw) Mosque.

It will realize that also through sending a number of its staff to specialize in what is known as "Area Studies" so that they may supervise the flow and modernization of information regarding the main areas of the world from which pilgrims arrive, the nature of the societal and political forces that are connected with the movement of pilgrims, the personalities that have an impact on them, the position of pilgrims in the stages pertaining to the adoption of political decisions in the countries concerned, its weight in relation to the ties of each country with the Kingdom, and the history of Hajj in those areas, in its cultural, literary, creative and religious dimensions [7].

## III. SOLUTION TO INFORMATION LEAKAGE

To manage an event like Hajj, organisers need to employ the best available technology to ensure the wellbeing of its participants. Current Hajj management system, based on partly computerised files, is insufficient to manage the large number of pilgrims. There should be a new Framework that uses sensor and RFID networks to track pilgrims during the extended period of Hajj. The new framework would store the data of the pilgrims within the central database, which would be linked to the sensor and RFID networks [8]. This would not only provide a framework for storing and retrieving pilgrim information but also be able to track and identify lost or dead pilgrims. As discussed, the current system is incapable to track, in real time, the pilgrims that are lost, injured or dead [9].

Most of the current management problems can be solved through integrating a Centralised Hajj Database together with suitable Sensor & RFID networks. In the proposed system, the Hajj database will be used for storing and retrieving the pilgrim information [10]. This with the Sensor network may be used to track or identify pilgrims within the designated Hajj areas.

The development of the Hajj Management System will require a detailed analysis of the current processes for managing the workflow of pilgrims during the pilgrimage of Hajj using different modelling techniques to determine requirements and be able to develop appropriate solutions [11]. This will enable the swift integration of technologies and the incorporation of any changes throughout the phases, such as the ability to track lost pilgrims during Hajj, if this was identified as a requirement.

## IV. RISKS FOUND IN HAJJ EVENT

- The current administration and management of pilgrims of Hajj is largely manual, except for using some computerised file systems. Some government departments use their own management systems but their information is not integrated with the current administration and management of the Hajj. Data is often stored in files of different formats for different departments. Many of these files use their own means of identification, which would make it difficult to someone wanting to link the individual databases. Since there is no centralised database system for the overall management of information, there are many problems when attempting to compile personal information, which is scattered across several incompatible filing systems.
- During the current registration process, pilgrims are provided with identifying wrist bands; these wrist tags provide only limited information about the pilgrim's identity. In cases where these bands are damaged or lost there can be major problems in identifying the

- pilgrim, especially when the pilgrim is lost and is unable to communicate. This may lead the identification to be impossible.
- The current system has no provision in place for collecting medical details of the pilgrims before issuing them with a visa. This puts other pilgrims at risk of catching communicable diseases, carried by affected pilgrims. Currently, each pilgrim is required to undergo certain immunisations before a visa can be granted. However, immunisations differ from country to country and no records are kept of individuals' immunisations. The pilgrims from the third world countries are most likely not to fully implement the immunisation process.
  - When so many people are gathered in one place it is inevitable that people will get sick and that there will be some deaths. In order to deal with such eminent situations, hospitals require medical and personal information about their patients. Hospital staff can have difficulty in identifying pilgrims especially those who come from other countries. In cases when a patient dies without leaving any visible information, the hospital may not be able to inform the Maulim about the fate of the pilgrim. Even if the wrist band was found with the body, it may take considerable time, sometimes up to days, to identify the body. In some cases, especially when the tag is missing, the body may never be identified. The police and the morgue suffer the same problem. They may also be unable to identify a pilgrim who has been arrested or has died.
  - The Maulim is responsible for managing all the pilgrims in their group and looking after them. The pilgrimage encompasses many places and the number of pilgrims usually is between four to six million. This exacerbates the problems of looking after a group of people. If a problem occurs with one of the pilgrims, that is if they are lost or injured, the Maulim may not be able to find them and may not know that they are in trouble; since tracking is impossible.
  - Passport handling is a big issue. When pilgrims reach at the airport, their passports are given to and held by the groups Maulim. From that time until the pilgrims return to the airport, at the end of pilgrimage, the passports remain in the custody of the Maulim. When pilgrims move from one city to another, the Maulim carries the passports and produces them at the various checkpoints along the journey. At the end of the journey, the passports are returned to the office of the Maulim. In this process, some passports can be lost, since they are being carried throughout the pilgrimage. Moreover, it can

be difficult to expedite journeys in emergencies.

- Hajj Fraudsters: Muslims shopping around for the best deal on a trip to Mecca, both in their local community and increasingly online, are attracted by packages flights, accommodation, visas which appear to offer good value for money. Some operators advertise large reductions. Individuals are asked to pay in cash or make a direct bank transfer prior to their trip and are told they will receive their tickets and travel documents nearer to the departure date. For some they never arrive.

## V. RISK ASSESSMENT IN HAJJ

Risk assessment is an important step in protecting people/workers in project, event, and business. For most, that means simple, cheap and effective measures to ensure your most valuable asset your workforce are protected. A risk assessment is simply a careful examination of what, in your work, could cause harm to people as complying with the law. It helps you focus on the risks that really matter in your workplace the ones with the potential to cause harm [13]. In many instances, straightforward measures can readily control risk, so that you can weigh up whether you have taken enough precautions or should do more to prevent harm. Workers and others have a right to be protected from harm caused by a failure to take reasonable control measures.

Accidents and ill health can ruin lives and affect your business too if output is lost, machinery is damaged, insurance costs increase or you have to go to court. You are legally required to assess the risks in your workplace so that you put in place a plan to control the risks. A risk assessment should be carried out for a proposed event considering all of the hazards, the nature and extent of the risks, and the action required to control them. This will be a legal requirement in many circumstances. There are 4 steps to completing the risk assessment form:

- Identify hazards
- Identify who could be harmed
- Assessing the risk
- Action to control the risk.

### A. *Identifying the hazards*

A hazard is something with the potential to cause harm. Examples of things that should be taken into account include:

- Any slipping, tripping or falling hazards
- Hazards relating to fire risks or fire evacuation procedures
- Any chemicals or other substances hazardous to health e.g. dust or fumes
- Any vehicles on site



- Electrical safety e.g. use of any portable electrical appliances
- Manual handling activities
- High noise levels
- Poor lighting, heating or ventilation
- Any possible risk from specific demonstrations or activities
- Crowd intensity and pinch points

#### B. Identifying those at risk

After hazard identification has been done, all groups of people who may be affected are specified for instance: Hajj managers, pilgrims, airport officers, immigration fellows, security people; those taking part in road users; members of the public (including children, elderly persons, expectant mothers, disabled persons), local residents and potential trespassers.

#### C. Assessing the risk

The extent of the risk arising from the hazards identified must be evaluated and the existing control measures taken into account. The risk is the likelihood of the harm arising from the hazard. For each hazard note down the severity number and the likelihood number using the Risk assessment Matrix see Fig. 1 below) [14]. This process will produce a risk rating of HIGH, MEDIUM, or LOW.

#### D. Action to control the risk

For each risk consider whether or not it can be eliminated completely. If it cannot, then decide what must be done to reduce it to an acceptable level. Only use personal protective equipment as a last resort when there is nothing else you can reasonably do. Consider the following:

- Remove the hazard
- Prevent access to the hazard e.g. by guarding dangerous parts of machinery
- Implement procedures to reduce exposure to the hazard
- The use of personal protective equipment
- Find a substitute for that activity/machine, etc.

The residual risk is the portion of risk remaining after control measures have been implemented. Fig. 2 [14] gives suggested actions for the three different levels of residual risk.

## VI. CONCLUSION

The Hajj event is unique in numerous respects, particularly in measures of scale and mass migration. It presents a challenge that impacts the international public risks as an increasing number of humans become more mobile, with everything this entails in terms of potential risks disease transmission and other health hazards.

Hajj management is an overwhelming task. International collaboration (in planning vaccination campaigns, developing visa quotas, arranging rapid repatriation, managing health hazards at the Hajj and providing care beyond the holy sites) is vital [15]. The most important role is assigned to the Saudi Arabia authorities, whose work and preparation for a mass gathering of such proportions is decisive and integral for the managing of the Hajj [16] and the outcome of the whole event.

## REFERENCES

- [1] Al-hashedi, A. H., Arshad, M. R., Mohamed, H. H., & Baharudin, A. S. (2012). RFID Adoption Intention in Hajj Organizations. *ICIT*, 386-391.
- [2] Bala Varanasi, U. &. (2012). A NOVEL APPROACH TO MANAGE INFORMATION SECURITY USING COBIT. *International Conference on Advances in Computing and Emerging E- Learning Techn.*
- [3] Clingingsmith, D., Khwaja, A. I., & Kremer, M. (2008). Estimating the Impact of the Hajj: Religion and Tolerance in Islam's Global Gathering.
- [4] Collins, N., & Murphy, J. (2010). THE HAJJ An Illustration of 360-Degree Authenticity. In *Tourism in the Muslim World* (pp. 321-340). Australia: Emerald.
- [5] Jo, H. &. (2011). Advanced Information Security Management Evaluation System. *KSH TRANSACTIONS ON INTERNET AND INFORMATION SYSTEMS*, 2221-4275, Vol. 05, No. 06, pp. .
- [6] L, J., & Brock, J. (2008). *Information Security Risk Assessment Practices of Leading Organizations*. US: GAO.
- [7] Revela, J. A. (n.d.). Data Leakage: Affordable data leakage Risk management. *Priviti*.
- [8] Rezakhani, A. &. & AbdolMajid & Mohammadi, N. (2011). (2011) "Standardization of all Information Security Management Systems. *International Journal of Computer Applications*, Vol.18, No. 8.
- [9] Sardar, Z. (2007). The Information Unit of the Hajj Research Centre. *Emerald*.
- [10] Sun, L., Srivastava, R. P., & Mock, T. J. (2006). An Information Systems Security Risk Assessment Model under Dempster-Shafer Theory of Belief Functions. *Journal of Management Information Systems*, Vol. 22, No. 4.
- [11] Toosarvandani, M. S., Modiri, N., & Afzali, M. (2012). THE RISK ASSESSMENT AND TREATMENT APPROACH IN ORDER TO PROVIDE LAN SECURITY BASED ON ISMS STANDARD. *International Journal in Foundations of Computer Science & Technology*, Vol. 2, No.6.
- [12] Whang, S. E., & Garcia-Molina, H. (n.d.). Managing Information Leakage. *Emerald*, 2-10.
- [13] Yamin, M. (n.d.). A FRAMEWORK FOR IMPROVED HAJJ MANAGEMENT AND RESEARCH.
- [14] Darlington. (n.d.). *Event Risk Assessment*. Environmental Health Section Town Hall Darlington.
- [15] Misra, S. C., Kumar, V., & Kumar, U. (2007). A strategic modeling technique for information security risk assessment. *Information Management & Computer Security*, 64 - 77.
- [16] Butt, M. (2011). Risk assessment. *Accounting, Auditing & Accountability Journal*, 131 - 131.



Risk Assessment Matrix (Probability and Likelihood Scales)							
Severity Rating	Description	Likelihood rating					
		1	2	3	4	5	6
		Very Unlikely	Unlikely	May happen	Likely	Very likely	Certain or imminent
1	Manual Administration and Management						
2	Minor injury, Loss of wrist bands		LOW				
3	Loss time Injury illness, major damage						
4	Major Injury, disabling illness, major damage			MEDIUM			
5	Haji Fraudsters						
6	Loss of Passport					HIGH	

Figure 1

Residual Risk	Action
Low Risk	No further improvements necessary provided control measures are in place and maintained. Continuous improvements should be sought during the review.
Medium Risk	Although risk is tolerable when control measures have been identified and implemented, further risk reduction measures are needed.
High Risk	Further Risk Reductio Measures <b>MUST</b> be undertaken.

Figure 2

## Performance Evaluation of DWDM TECHNOLOGY: An Overview

Shaista Rais<sup>1</sup>, Dr.Sadiq Ali Khan<sup>2</sup>

Department of Computer Science, University of Karachi.

### ABSTRACT:

For abstract we shall discuss the following method. DWDM: dense wavelength division multiplexing. It is the method for expanding the data transfer capacity of optical system interchanges. DWDM controls wavelength of light to keep sign inside its own specific light band. In DWDM system dispersion and optical sign are the key elements. Raman and 100G advances are particularly discussed.

### 1.INTRODUCTION:

DWDM allows many particular and specific information science over a solitary fiber. A fiber optic link inside helps light from end to end. This is a sign infused in one end by a LED. (light-discharging diode) by lasers. Leds can create flags up to around 300 Mbits/sec. Lasers can produce motions in the multi-gigabit/sec. Leds are used for short-separate optical connections. Lasers deliver light in “windows”. A window is an infrared range that is developed for optical transmission. The ITU occupies groups for fiber optic frameworks. Fiber optics uses light to transmit signals. An optical fiber has three segments:

- i. The center (core) conveys the light signals.
- ii. The refractive index contrast between center and cladding limits the light to the center (core).
- iii. The covering (coating) shields the glass.

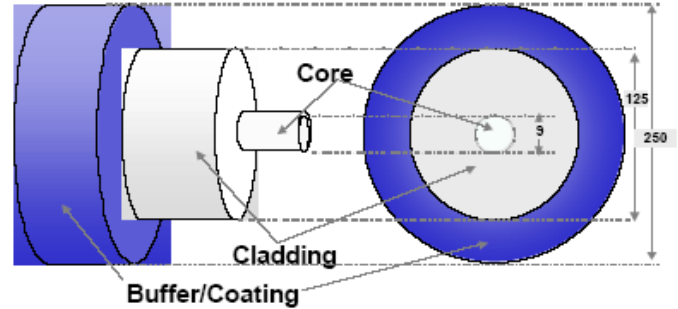


Figure 1. Structure of Optical Fiber

Fiber optic frameworks utilize light (signals) inside the infrared band (1mm to 400 nm wavelengths) of the electromagnetic range. Frequencies of light in the optical range of the electromagnetic range will be normally recognized by their wavelength, despite the fact that recurrence (distance between lambdas) gives a more particular known proof.

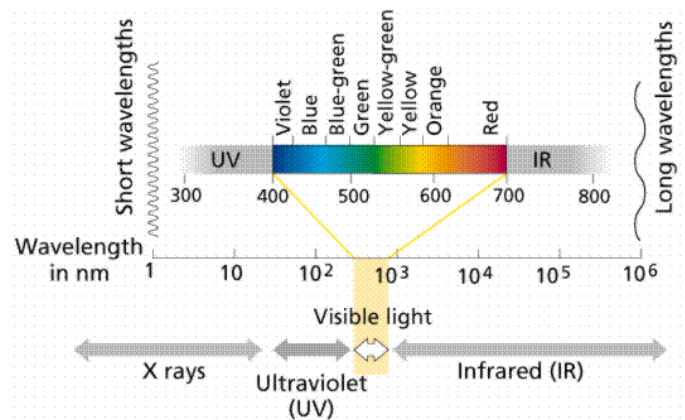


Figure 2: Optical communication wavelength bands in the infrared.

Wavelength multiplexing method utilizes the way that light can be going down the fiber all the while with various colors of

light without blending or meddling with one another.

## 2..DWDM COMPONENTS:

- i. Transmitter/ Receiver
- ii. Multiplexer/ Demultiplexer
- iii. Optical Add/Drop Multiplexer
- iv. Optical Amplifiers
- v. Transponder

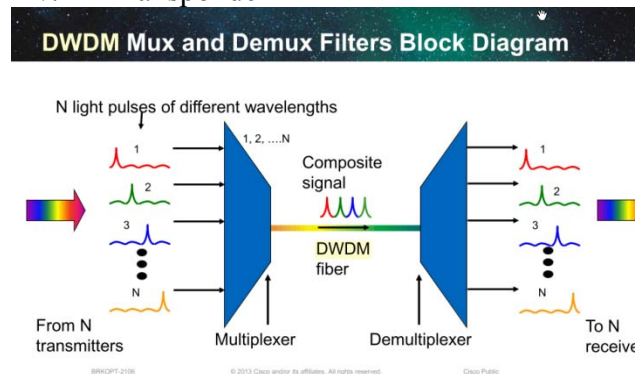


Figure 3: Block diagram of Fiber Optic Transmission system

### Details of the above components:

#### 2.1.Transmitter / Receiver

Fiber optic transmission gives the same essential components as copper-based transmission frameworks. A transmitter, (receiver), and a medium by which the sign is passed from one to the next. For this situation, optical fiber as shown in figure 3.

#### 2.2.Multiplexer / Demultiplexer

- It is the methodology in which various data streams starting from opposite sources are consolidated and transmitted over a single data. Multiplexing is followed by a gear called multiplexing (MUX).
- It is put at the Transmitting End of the correspondence join. At the Receiving End, the Composite Signal is differentiated by a gear called Demultiplexer (DEMUX).

Demultiplexer performs the opposite methodology of Multiplexing and courses the divided signs to their comparing Receivers or Destinations shown in figure 3.

#### 2.3.Optical Add/Drop Multiplexer

- It is used as a part of wavelength division multiplexing steering fiber optic signs. The include and drop individual sets of wavelength channels from a thick wavelength division multiplexing (DWDM) multi channel.

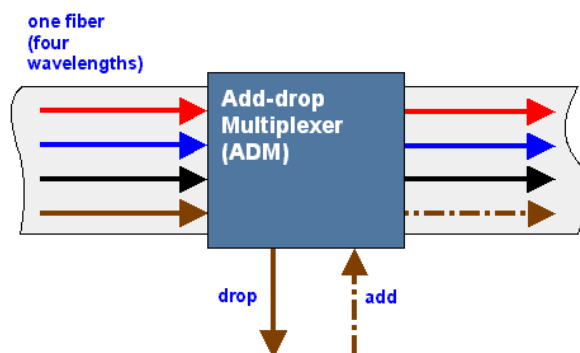


Figure 4: Optical Add/drop Multiplexer

#### 2.4. Optical Amplifiers:

- It is an essential optical correspondence connection including a transmitter and beneficiary, with an optical fiber link uniting them.
- Mediums, for example, copper, there is still a cutoff of around 100 km on the separation, the signs can go before getting to be so loud, it would be impossible to be located.
- Before the commercialization of optical speakers, it was necessary to

electronically recover the optical flags each 80-100 km with a particular end goal to complete transmission over long separations.

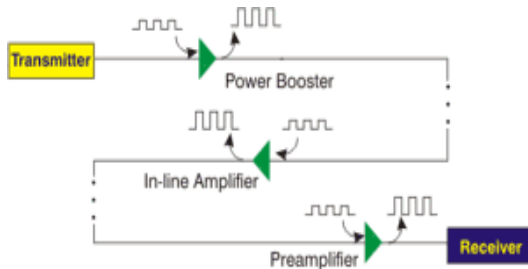


Figure 4: Optical Amplifier

This implied getting the optical flag, cleaning and enhancing it electronically, and after that retransmitting it throughout the following section of the correspondence join.

### 2.5. Transponder (Wavelength Converter)

- transponders change over optical signals from one nearing wavelength to an alternate cordial wavelength suitable for DWDM applications.

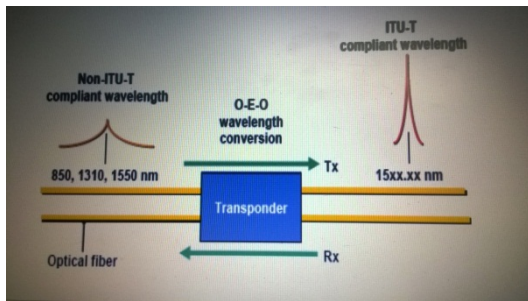


Figure 5: Transponder Performance

Transponder performance in four steps:

- The transponder acknowledges info in the structure of a standard single-mode or multimode laser beat.
- The info can originate from diverse physical media and different conventions and activity sorts.

- The wavelength of the transponder info sign is on DWDM wavelength.
- DWDM wavelengths from the transponder will be multiplexed from the immediate interface to structure a composite optical signal which will be sent.

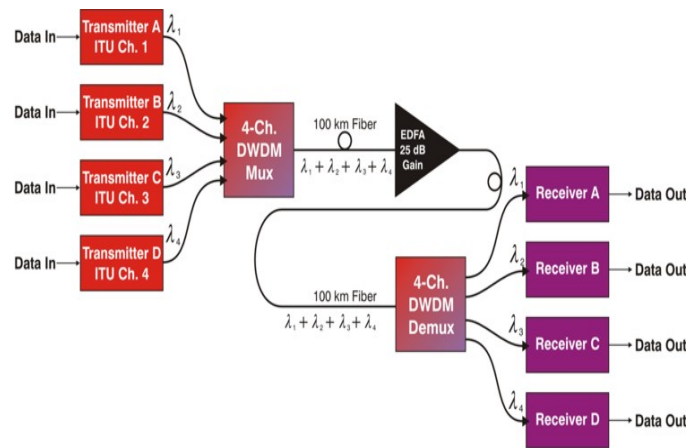


Figure 6: Block Diagram of How Transponder works

Sending DWDM frameworks permitted transporters to serve tremendous limit of information via convey different wavelengths over a solitary pair of fiber. For transporters, this implies a noteworthy expense investment funds contrasted with the expense of sending single-channel systems or overlaying different systems for each one administration advertising. Notwithstanding, because of increment in web activity particularly feature movement and versatile backhaul, even 40-channel WDM frameworks worked at 10 Gbps every wavelength are rapidly arriving at limit on intensely utilized courses. Transporters require a financially savvy answer for growing limit, while in the meantime diminishing the expense every bit transported. To meet quick movement development, numerous merchants have created 40 Gbps transponders and muxponders that can be utilized with existing sent WDM systems offering a 4×

increment in system limit. Its genuine test for bearers to get the transmission information rate increment by utilizing the current framework. While 40g offers quick advantages, continually expanding information development (Youtube, cell phone and so on.) obliges sending of considerably higher-limit DWDM frameworks

(e.g. 80 channels) with higher information rates (100 Gbps) used on every wavelength. Lamentably, as optical velocities expand, it gets to be progressively hard to overcome optical hindrances as velocity is specifically proportionate to the disability and still attain to satisfactory execution. The business is beginning to tackle these 100g transmission issues and create key 100g optical parts expected to actualize and send 100g in true systems.

#### Engineering Issues

Optical flags in a fiber are presented to various ruinous optical hindrances including Chromatic Dispersion (CD) and Polarization Mode Dispersion (PMD). These impedances bend the honesty of the first optical signs; constraining the separation they can be transported.

Upto 10gbps information arte, the industry has discovered strategies to tackle and make up for these optical debilitations. Sadly, the optical debilitations deteriorate and significantly more hard to make up for as rates increment, as demonstrated in Figure 7. Creative procedures are obliged to transmit 100g over metro, provincial and whole deal optical systems. We can see the debilitation gets intensify by 10 times when we bounce

from 10gbps to 100gbp.

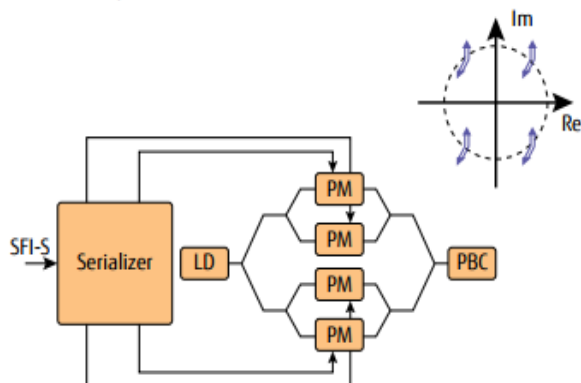
Transmission Impairment Sensitivity			
	10 Gbps	40 Gbps	100 Gbps
CD sensitivity	1	x 16	x 100
PMD sensitivity	1	x 4	x 10
ONSR requirement	1	+ 6 dB	+ 10 dB
Optical band-pass sensitivity	1	x 4	x 10
Sensitivity to fiber nonlinearity	1	x 4	x 10

Figure 7: Transmission Impairment Sensitivity

#### Modulation

Adjustment – pushing more bits immediately Given that optical execution weaknesses intensify with expanding speeds, one approach to take care of this issue is to send various bits of information down the fiber all the while, lessening the general image rate. Encoding different bits of data into an image is known as "tweak." Up to 10 Gbps, optical frameworks use straightforward on-off keying (OOK) to speak to the advanced 1s and 0s. A "1" is the laser light turned on, and a "0" is the laser light turned off, so one bit of data is transmitted with every optical image. This is an extremely basic and financially savvy tweak procedure, and it has worked exceedingly well in optical systems up to 10 Gbps. At higher information rates, more advanced regulation procedures are obliged to minimize the impacts of optical debilitations. At 100g, the industry has institutionalized on a balance plan known as DP-QPSK, as indicated in Figure 2. A DP-QPSK modulator is generally mind boggling and immoderate to actualize, yet it permits four bits of information to be encoded and sent as one optical image. With the lower optical image rate, 25 Gbaud rather than 100 Gbps, optical hindrances are sort of simpler to adjust.

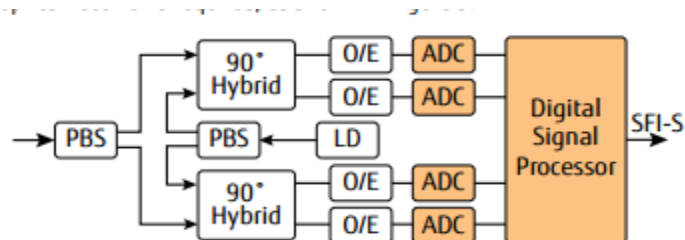




**Figure 8:100G Modulator based on DP-QPSK**

### Coherent Receivers

Rational recipients – recouping the bits Like 100g transmitters, 100g optical beneficiaries are more intricate because of the complex DP-QPSK balance plan. Up to 10 Gbps, a straightforward photograph finder changed over the approaching photons to computerized "1" or "0" flags, an extremely basic and moderately minimal effort approach. With the presentation of 100g DP-QPSK tweak, a substantially more perplexing optical beneficiary is needed, as indicated in Figure 3.the extra intricacy has a couple of profits. The rapid Adcs and DSP permit optical disabilities, for example, CD and PMD, to be repaid electronically inside the DSP, offering a bigger scope of recompense and better control. The DSP empowers various extra recompense and estimation strategies to be executed that essentially weren't conceivable with the optical-just photograph locator beneficiaries utilized at 10 Gbps and lower rates. For most merchants, these calculations are the genuine mystery sauce of their 100g optical units, and they separate the execution levels of these units contrasted with their competitors.



### 3.100G application

At a very high state, the voracious interest for rapid information administrations, especially Internet feature administrations, is driving the requirement for higher system limits. Inside optical transport organizes; this interest for extra transmission capacity is bringing about a few essential 100g applications:

- Backbone capacity expansion
- Router Interconnect

Spine (Backbone) limit development Wavelengths over a LH or ULH a piece of the system are at a premium, because of the cost of the long term wavelengths and in addition the separations voyaged. For instance, including overlay limit along a Chicago-to-Atlanta course can be restrictive because of the long separations and the quantity of OADM and ILA hubs needed. Far more atrocious is a circumstance in which a bearer is out of limit and out of extra strands along the course, following the main option is an extremely costly extend to burrow, force and develop new fiber along the whole course – effortlessly a multimillion-dollar venture To guarantee these wavelengths are completely used, most metro territories have conglomeration stages that consolidate lower-rate administrations into huge, full 10g funnels for hand-off to

the spine system. Because of the expense of the whole deal wavelengths, redesigning spine systems will be one of the first applications for 100g muxponders and transponders. By including 100g muxponders at the total focuses, bearers can grow their spine limits by a variable of 10 without the expense of extra WDM overlay systems or fiber development.

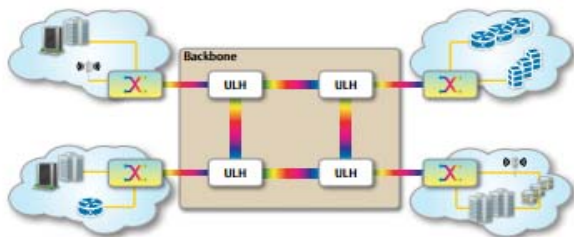


Figure 6: 100G backbone capacity expansion

### Router Interconnect

The optional 100g application is for switch interconnection as indicated in Figure 10. Today, numerous switches use different 10g associations over a system, as indicated on the left half of the outline. As the quantity of 10g switch interconnections expands, switch expenses go up, switch proficiency decreases and operational intricacy increments for the transporter. At one point, it bodes well for switch to a solitary switch port running at 100g. As 100g switch interfaces get to be accessible, transporters will begin to see more demands for 100g private-line administration to interconnect these fast switches. The switch interconnect application happens crosswise over metro systems, territorial systems and national spine systems.



### ISSUES IN CABLE PERFORMANCE:

Fiber optic link has specific qualities that breaking point its execution. Assembling of the link from distinctive producers may display varieties in these characteristics; Attenuation and Dispersion are the two primary issues that influence the essential execution of the fiber transmission framework.

Three sorts of gadgets may be utilized to overcome constriction:

- **Electronic Regenerator:** This gadget recovers motions by first changing over optical signs to electrical signs. The electrical sign is recovered, changed over once again to optical, and infused go into the fiber. Regenerators are excessively wasteful for current high velocity optical systems because of electrical recovery necessities. On WDM frameworks, every wavelength obliges its own particular opto-electric enhancer, an extravagant suggestion if there are numerous wavelengths.
- **EDFA** Short for erbium-doped fiber intensifier. EDFA is an optical repeater gadget that is utilized to support the force of optical signs being brought through a fiber optic correspondences framework. An optical fiber is doped with the uncommon earth component erbium so the glass fiber can ingest light at one recurrence and radiate light at an alternate recurrence. An outer semiconductor laser couples light into the fiber at infrared wavelengths of either 980 or



1480 nanometers. This activity energizes the erbium ions. Extra optical signs at wavelengths somewhere around 1530 and 1620 nanometers enter the fiber and fortify the energized erbium molecules to discharge photons at the same wavelength as the approaching sign. This activity opens up a feeble optical sign to a higher force, influencing a support in the signal strength.

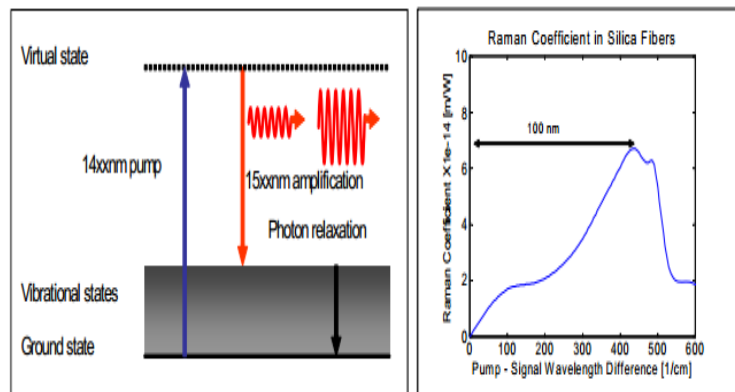
#### 4. RAMAN Amplifier:

RAMAN enhancer is an alternate critical speaker innovation other than EDFA. The working foremost behind Raman speaker is called animated Raman dissipating which was found by Sir Chandrasekhar Raman in 1928. He portrayed a methodology when light photons collaborate with matter atoms; they disperse to a higher wavelength. The photon energizes the matter atoms to a high (virtual) vitality state, which then unwinds once again to the starting state by radiating an alternate photon and also vibrational vitality. Because of the vibrational vitality, the transmitted photon has less vitality than the occurrence photon, and consequently a higher wavelength.

Keeping in mind the end goal to get fitting optical enhancement we need animated Raman dissipating ( a nonlinear impact in optical fiber) which is a comparative methodology when a higher wavelength photon energize the scrambling procedure, i.e. the ingestion of the introductory photon, bringing about the discharge of a second higher wavelength photon, therefore giving intensification. This is indicated in Figure 1 for silica strands, where a ~1550nm sign is opened up through retention of pump vitality at ~1450nm.

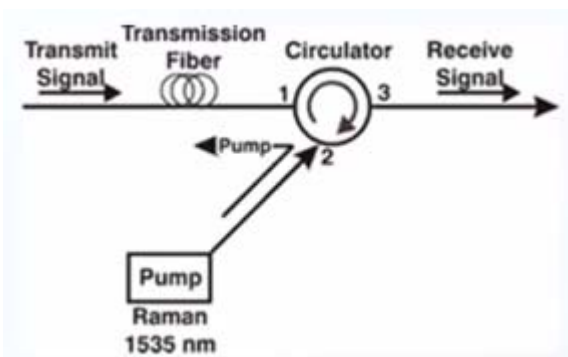
In Raman intensification the increase range relies on upon the wavelength of pump with greatest addition happens around 100nm

higher than pump wavelength. Then again in EDFAs, the increase ranges is steady and focus by Erbium ions.

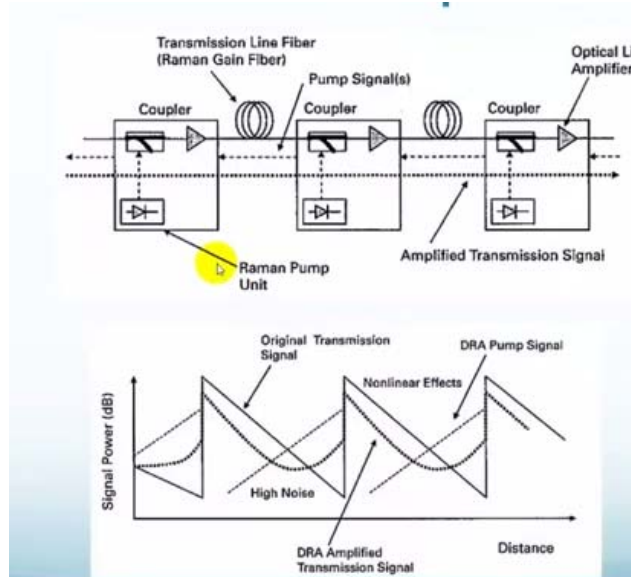


There are 2 types of Raman amplifiers:

- Dispersed Raman Amplifiers: Dispersed Raman Amplifiers or DRA which utilizes the transmission itself as a medium into which a high power pump laser (Raman pump) is infused at the far end which flies out regressive to increase the signal.
- Discrete Raman Amplifier: In Discrete Raman Amplifier, a loop of committed fiber is utilized alongside a pump laser.



In real world, Raman Amplifiers are utilized together with EDFA intensifier to completely use their separate playing point.



Raman intensifier, open up the sign in retrogressive course and EDFA increase the sign in forward bearing. The comparing sign force level is demonstrated in the base picture, the first flag (spotted line) first gets weaker by fiber misfortune yet later it step by step increments by infusing the retrogressive Raman pump and directly after the Raman speaker, the sign is instantly supported again by EDFA enhancer(amplifier).

### Dispersion/Scattering

An alternate trademark is scattering, which is the expanding of a light heartbeat as it goes down the link. Exorbitant scattering will make a sign hard to peruse by the collector. At the point when a LED or laser sends light into a multimode fiber, a scope of wavelengths of light is available. Some of those wavelengths go at distinctive rates than others. The impact is to misshape waveforms, which can result in blunders in perusing the sign at the flip side of the link. Reviewed record link is intended to

minimize the postponement of the slower wavelengths.

There are four types of dispersion:

- Material scattering Variations in the refractive properties of the link reason signal scattering.
- Modal dispersion Occurs in multimode link. Light takes diverse ways through the link with light on a few ways having a more drawn out travel time than others. Reviewed link adjusts this effect.
- Chromatic dispersion: This happens in light of the fact that a few wavelengths go through a medium speedier than others. The more drawn out the link, the more regrettable the impact, and the harder it is to peruse the sign.
- Waveguide dispersion: This scattering happens in single-mode fiber because of the distinction in the velocity of the sign between the center and the cladding. It causes chromatic scattering.
- G.652 link, utilized in most business frameworks, exploits the 1,310-nm window where chromatic scattering is minimized. This window is regularly called the zero-scattering point-it is the extent where chromatic scattering is minimized on the grounds that the waveguide scattering offsets material scattering.
- Long-pull bearers, then again, have higher data transmission and separation prerequisites, so G.653 and G.655 "scattering moved" fiber working in the C band is favored. The C band is utilized for DWDM frameworks, which help numerous

nearly divided channels at information rates of 10 Gbits/sec and higher. Two different groups are presently being utilized to help limit and separation: the 1,460 nm to 1,530 nm S band and the 1565 to 1625 L band. A more up to date methodology utilizes solution innovation, which can be utilized to make a link framework that extends almost partially far and wide.

- Note that it is conceivable to help DWDM on numerous more seasoned fiber links. Standard single-mode fiber will help DWDM at lower information rates. Some more seasoned scattering moved filaments were not ready to handle DWDM, however these links may be made to act like nonzero scattering link by utilizing wavelength above and underneath the 1,550-nm window.
- Corning and Lucent are the real suppliers of whole deal link. Lucent's Truewave and Allwave links are made of single-mode nonzero scattering strands that help all the wavelength windows. True wave is particularly intended for optically intensified, high-powered long-remove DWDM systems working in both the C band and the L band. Both link sorts are produced with a protected purging methodology to uproot water particles in the center, along these lines permitting more extensive range utilization.
- Corning's LEAF is a solitary mode NZ-DSF fiber intended for DWDM frameworks. It joins low lessening and low scattering with a powerful range that is 32 percent bigger than non- NZ-DSF fiber. This permits more power to be pumped into the system over more channels without nonlinear impacts that make clamor,

contort flags, and debase execution. It can work at 10 Gbits/sec or higher utilizing high-yield power EDFAs.

- Corning's Metrocor fiber is a solitary mode NZ-DSF link enhanced for short-separate metropolitan use. It doesn't require the effective lasers that are needed in the whole deal environment thus aides lessen the expense of actualizing metropolitan fiber systems.

Bearer systems are confronting huge increments in transfer speed transported, because of a blend of higher-pace business information administrations, portable PDAs and Internet feature activity. In the meantime, these systems are under huge weight to bring down their expense every bit transported. The arrangement is to convey more bits every wavelength by using 100g, ensuing in better system usage, higher ghostly efficiencies and lower general working expenses. Higher optical velocities, past 10 Gbps, bring about huge execution punishments because of optical impedances, for example, chromatic scattering and PMD.

Propelled tweak procedures and cognizant recipients defeat these optical debilitations and confinements, permitting 100g transponders and muxponders to be conveyed over existing systems. In 2010, the optical segment industry reacted with 100g coordinated optical segments, which empower the presentation of 100g transponders and muxponders in 2011. Beginning 100g applications will probably be centered around expanding limit crosswise over spine systems and 100g switch interconnections.

## 5.COMPARISON OF TRANSMISSION PERFORMANCE OF THREE OPTICAL FIBERS:

We think about the transmission execution of three diverse optical filaments in discrete

256 Gb/s PM-16qam frameworks intensified with erbium doped fiber amplifiers (EDFAs) and disseminated Raman enhancement. The compass length in every framework is 100 km. The filaments contemplated incorporate standard single-mode fiber, single-mode fiber with ultra-low misfortune, and ultra-low misfortune fiber with substantial successful territory. We find that the single mode fiber with ultra-low misfortune and the huge viable range fiber with ultralow misfortune manage the cost of achieve favorable circumstances of upto around 31% and 80%, individually, over standard fiber measured at separations with 3 db edge over the forward slip revision (FEC) limit. The Raman intensified frameworks give around half achieve length upgrade over the EDFA frameworks for each of the three filaments in the test set-up. For the best performing fiber with expansive successful zone and ultra-low misfortune, indisputably the achieve lengths with 3 db edge are more noteworthy than 1140 km and 1700 km for the for EDFA and Raman frameworks, separately.

## INTRODUCTION TO TRANSMISSION PERFORMANCE

As overall information activity keeps on growing, comparative development necessities are put on the limit of whole deal fiber-optic correspondence. This basic has driven scientists to explore signal bit rates and otherworldly efficiencies past 100 Gb/s, 2.0 b/s/Hz in sound optical frameworks. The following venture up in limit without expanding the framework image rate might be 200 Gb/s every wavelength channel utilizing 16-ary quadrature with excess balance (16qam), which gives 4 bits/image every polarization. Thusly, there has been a critical level of exertion as of late steered towards exhibiting 200 Gb/s every channel

with polarization multiplexed 16qam (PM-16qam) signs .

In any case, an essential test of this adjustment arrangement is its altogether decreased scope

in correlation to 100 Gb/s PM-QPSK signals. The lessened span is mostly attributable to the higher optical-signal-to-noise (OSNR) needed for 16qam signs than QPSK signals with the same image rate to attain to a similar bit mistake degree (BER). For perfect signs,

16qam signs require roughly 7 db higher OSNR than QPSK [7]. This would prompt an achieve lessening for 200 Gb/s PM-16qam by a component of around 5x in examination to 100 Gb/s PM-QPSK. By and by, nonetheless, the lessening in achieve may be more serious [5] because of bigger usage punishments connected with 16qam transmitters. In view of the higher OSNR prerequisites for 200 Gb/s PM-16qam frameworks, it is imperative to increment framework OSNR through every methods accessible. An essential course to upgrading OSNR and along these lines

expanding framework range is by the utilization of optical fiber with lower lessening and/or bigger

successful region. Both credits can serve to expand OSNR in examination to standard fiber. In this work, we specifically analyze tentatively measured transmission reach for 256 Gb/s PM-16qam frameworks in excess of three distinctive optical strands with both EDFA and Raman enhancement. The three strands incorporate standard G.652-agreeable single-mode fiber, G.652-consistent fiber with ultra-low misfortune, and a G.654-consistent fiber initially created for submarine applications with both ultra-low misfortune and a bigger successful region. We find that the extensive viable region, ultra-low misfortune fiber shows an

achieve advantage over the standard single-mode fiber of up to 80%. The achieve point of interest of the Raman frameworks was around half over the EDFA frameworks for all filaments in the test design.

#### EXPERIMENTAL SETUP:

A schematic chart of the general test set-up is demonstrated in Fig. 1. 20 optical wavelengths were balanced together with a 32 Gbaud 16qam transmitter. The channel under test was encoded on an outer cavity laser (ECL) with 100 khz linewidth, and the other 19 channels were encoded on DFB lasers. The channel dispersing was 50 Ghz. The channels were polarization multiplexed and afterward passed through a short bit of fiber (around 8 km, 160 ps/nm scattering) to de-relate nearby channels by  $> 2$  images before being propelled into a re-circling circle assembled with the given fiber under test (FUT). The circle transmission medium was contained three 100 km compasses of fiber with either a solitary stage EDFA or regressive pumped Raman intensifier toward the end of each one compass. On account of the Raman intensifiers, the pump flows were changed in accordance with give straightforwardness, i.e. Raman increase sufficient to adjust for the previous compass misfortune. Three pump wavelengths at 1427 nm, 1443 nm, and 1462 nm were utilized. The rest of the circle incorporated a circle synchronous polarization scrambler, an increase leveling filter, and an alternate EDFA to adjust for the circle component misfortune.

Channels were chosen for discovery in a polarization- and stage assorted computerized intelligible collector with a free-running neighborhood oscillator with 100 khz linewidth. The four signs from the adjusted photodetectors were digitized by simple to-computerized converters working

at 50 Gsamples/s utilizing an ongoing inspecting oscilloscope with 20 Ghz electrical data transfer capacity. BER qualities were computed through immediate mistake numbering over more than 3 million bits with logged off computerized sign transforming steps including quadrature lopsidedness recompense, up-inspecting to 64 Gsamples/s, chromatic scattering payment utilizing a recurrence area equalizer, advanced square-and-channel clock recuperation, polarization demultiplexing and adjustment utilizing a 21-tap versatile butterfly structure with channel coefficients dictated by a range regulated consistent modulus calculation (CMA) [8] taking after preconvergence utilizing a standard CMA, bearer recurrence counterbalance utilizing a ghastrly space calculation, and stage recuperation utilizing a food forward calculation [9]. All chromatic scattering was repaid digitally in the logged off handling.

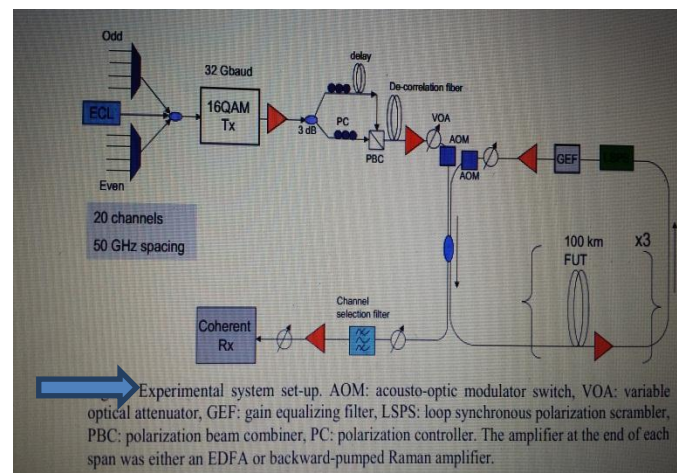


Figure 13: Experimental System Setup

The 16qam optical transmitter is demonstrated schematically in Fig. 2. Four double 32 Gbaud signs yield from a heartbeat design generator were consolidated in sets with 6 db relative lessening contrast between the signs in each one sets. Relative deferrals of a few several images served to de-associate the individual



2 15-1 PRBS signals from one another, and additionally the ensuing pair of 4-level heartbeat plentifulness balance (PAM) electrical signs driving an IQ optical modulator. The 32 Gbaud image rate delivers a general bit rate of 256 Gb/s for the PM-16qam signs, with a 28% aggregate overhead over the ostensible 200 Gb/s information rate.

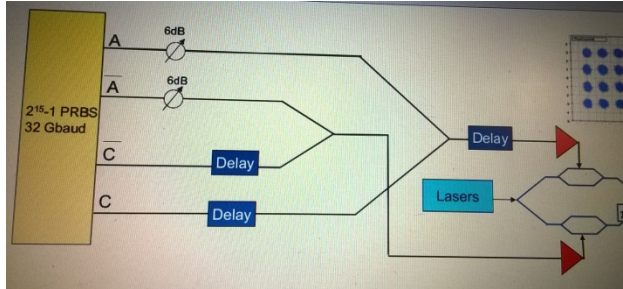


Figure 14: Simplified schematic diagram of the 16QAM Transmitter Configuration

The expected delicate choice FEC (SD-FEC) has a crude BER limit of  $2 \times 10^{-2}$  for a post-FEC lapse rate of  $< 10^{-15}$ . [10]

The three fibers tested were Corning  $\text{SMF-28e}^+$ ,  $\text{SMF-28}$ , ULL and Vascade  $\text{EX2000}$  optical fibers. The attenuation values for these 3 fibers were 0.191 dB/km, 0.166 dB/km, and 0.161 dB/km, respectively. The fiber effective areas were  $82 \mu\text{m}^2$ ,  $82 \mu\text{m}^2$  and  $112 \mu\text{m}^2$ , respectively. The nominal chromatic dispersion at 1550 nm is about 16.5 ps/nm/km for  $\text{SMF-28e}^+$  fiber, 16 ps/nm/km for  $\text{SMF-28}$  ULL fiber, and about 20 ps/nm/km for Vascade  $\text{EX2000}$  fiber.

## EXPERIMENTAL TRANSMISSION RESULTS:

A solitary channel consecutive OSNR affectability characterization of the PM-16qam transmitter and beneficiary was initially performed before transmission tests over fiber. The results are demonstrated in Fig. 3. When contrasted with a perfect

transmitter and recipient match, the exploratory usage punishment was around 3.2 db at a BER estimation of  $1 \times 10^{-3}$ .

Transmission tests over the three diverse optical strands were initially performed for the EDFA frameworks. The EDFA toward the end of each of the initial two compasses was a solitary stage EDFA. The EDFA toward the end of the third compass was a 2-stage preamp and supporter with the circle synchronous polarization scrambler and tunable addition evening out channel put in the

mid-stage. The same EDFAs were utilized for every one of the 3 fiber frameworks. For every fiber tried, we initially decided the ideal dispatch power into the compasses by measuring BER as a capacity of channel force at the separation of 600 km (2 circle disseminations). The results for the channel power streamlining for every one of the three strands are indicated in Fig. 4(a). The ideal channel force was around 0 dbm for both  $\text{SMF-28e}^+$  and Vascade  $\text{EX2000}$  strands, while it was about -1 dbm for  $\text{SMF-28}$  ULL fiber. The ideal dispatch force is dictated by both the fiber powerful territory and in addition the fiber attenuation.

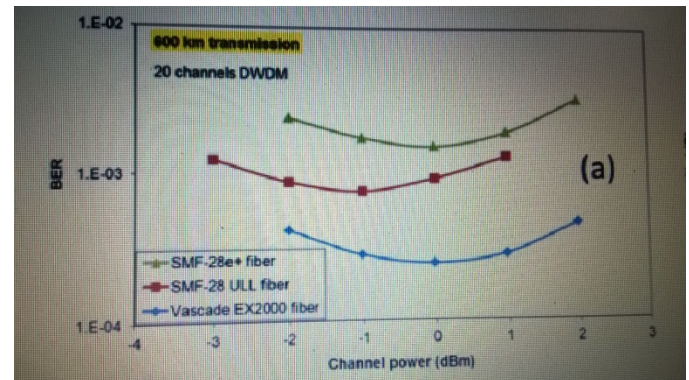


Figure 15(a) : Channel Power (dBm)

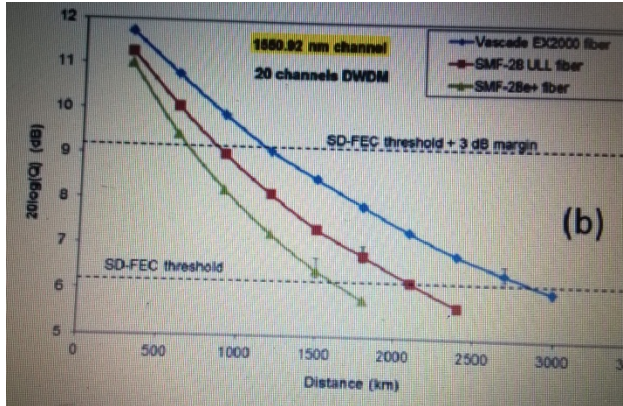


Figure 15 (b) : Distance (km)

With the ideal dispatch force set for every fiber framework, we then measured the BER of the 1550.92 nm direct in the middle of the 20 channel framework as a capacity of transmission separation. The consequences of these estimations are demonstrated in Fig. 15(b) regarding  $20\log(Q)$  in db as a capacity of separation.  $Q$  is computed by  $Q = \sqrt{2\text{erfc}^{-1}(2 \cdot \text{ber})}$ . Figure 15(b) additionally demonstrates two dashed lines relating to  $Q$  estimations of 6.25 db (the SD-FEC edge), and 9.25 db (3 db edge over the limit). The longest transmission scope is unmistakably gotten with the ultra-low misfortune and huge viable region Vascade Ex2000 fiber. Each of the 20 channels were measured with  $Q$  values over the FEC edge at 2700 km with this fiber. The spread of the  $Q$  values over each of the 20 channels is spoken to by the blunder bars in  $Q$  at this separation. Likewise, each of the 20 channels were additionally measured for the other two strands at the greatest circle separations underpinned and the  $Q$  spread in each one case is spoken to by slip bars. For every one of the three strands, the OSNR of the estimation channel for the separation with  $Q$  closest to the SD-FEC edge was around 22 db.

Exploratory transmission set-ups frequently incorporate a different modulator for odd and even channels as a method for guaranteeing that nearby channels have de-

related information streams. Our set-up utilized one modulator for all channels because of equipment confinements yet incorporated a bit of optical fiber on the transmitter side which served to de-associate adjoining channels by a couple images before dispatch into the re-circling circle through chromatic scattering. To check this present course of action's viability, we additionally measured the BER of the focal 1550.92 nm estimation channel with the two adjoining diverts turned off as a capacity of separation to contrast with the information with the nearby channels on. A case of the consequences of this correlation for the SMF-28 ULL fiber are indicated in Fig. 5 in which  $Q$  is plotted versus OSNR, with every information point speaking to an alternate separation. The almost insignificant contrast between the cases with neighboring diverts on and off proposes that the single modulator utilized in the transmitter does not fundamentally influence the nonlinear conduct of the framework with all scattering recompense performed digitally in the beneficiary.

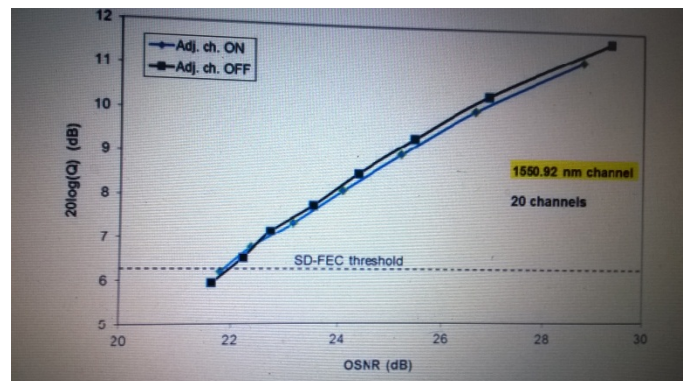


Figure 16:  $Q$  vs. OSNR for SMF-28 ULL fiber system with EDFAs

The transmission consequences of the EDFA frameworks are outlined in Figs. 6(a) and 6(b) as the achieve length for every fiber for shifting levels of  $Q$  edge over the SD-FE limit. The compass was dead set for the 1550.92 nm channel by interjection of the measured information focuses taken for



different circle disseminations. The achieve lengths in km are given in Fig. 6(a) while Fig. 6(b) demonstrates the achieve qualities standardized to that of the standard single-mode fiber. At the essential 3 db edge level, the Vascade Ex2000 fiber and the SMF-28 ULL fiber have achieved favorable circumstances over the standard single-mode fiber of 80% and 31%, separately. These relative results for SMF-28 ULL fiber contrasted with standard single-mode fiber are reliable with prior results acquired for 40 Gb/s non-intelligible transmission frameworks [11].

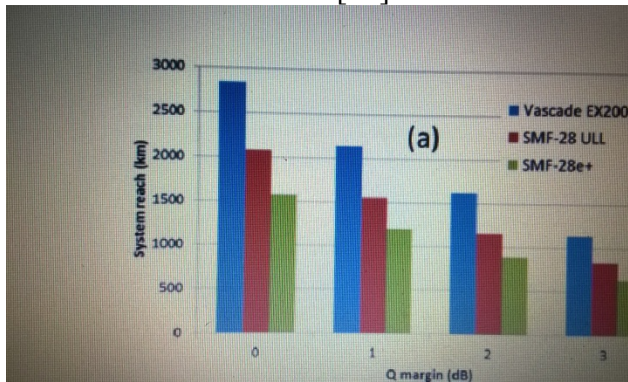


Figure 17(a): Summary of reach length results for EDFA systems with 100 km spans. (a) Absolute reach lengths in km.

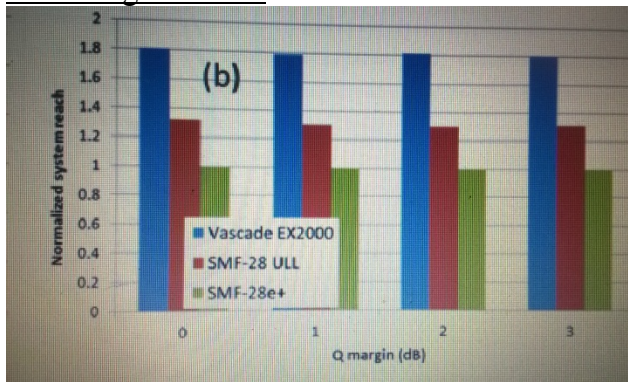


Figure 17(b): Reach lengths normalized to that of standard single-mode fiber.

Frameworks utilizing these same three optical filaments increased by retrogressive Raman pumping were explored next. For these frameworks, the Raman pump forces were balanced in

each case to completely make up for the loss of the first compass and the discrete loss of the Raman module. The evaluated aggregate Raman pump power every compass required was around 850, 750, and 935 mw for the SMF-28e + , SMF-28 ULL, and Vascade Ex2000 strands, separately. The promoter EDFA was utilized toward the end of the re-coursing circle to make up for the loss of the circle components. As in the recent past, ideal channel dispatch forces were initially decided, this time at 900 km separation, and after that the BER estimations of the focal point 1550.92 nm channel inside the 20 channel framework were measured at the optimal channel control as a capacity of transmission separation for all fiber frameworks. The results from these two sets of estimations are demonstrated in Figs. 7(a) and 7(b), individually. The ideal channel forces were 2.5-3 db lower for every fiber contrasted with the EDFA frameworks. The mistake bars in Fig. 7(b) again speak to the spread in Q qualities measured over each of the 20 channels at the separations at which all channels were over the SD-FEC limit. For the Raman frameworks, these separations were 2100 km, 2700 km, and 3600 km for SMF-28e + , SMF-28 ULL, and Vascade Ex2000, separately. At the SD-FEC edge, the OSNR estimations of the estimation channel were around 0.3-0.7 db higher than in the EDFA frameworks, maybe reflecting little punishments from twofold Rayleigh back-diffusing (DRBS) for the Raman frameworks. The framework built with Vascade Ex2000 had the littles such punishment because of the fiber's bigger successful range and littler obliged increase.



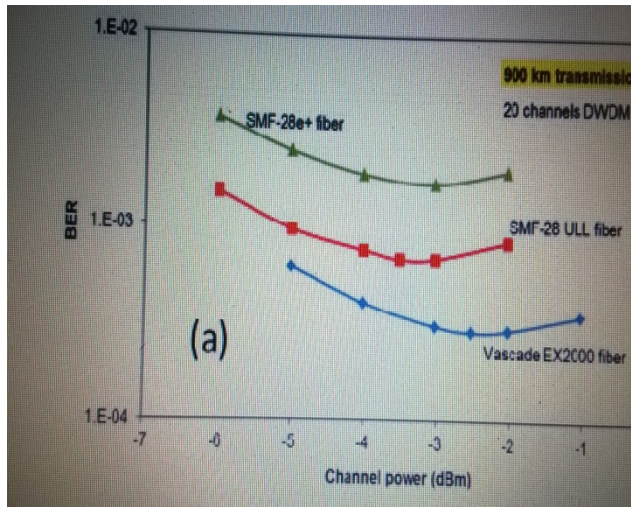


Figure 18(a): Raman amplified systems: (a) BER as a function of channel power for 1550.92 nm channel in 20 channel systems.

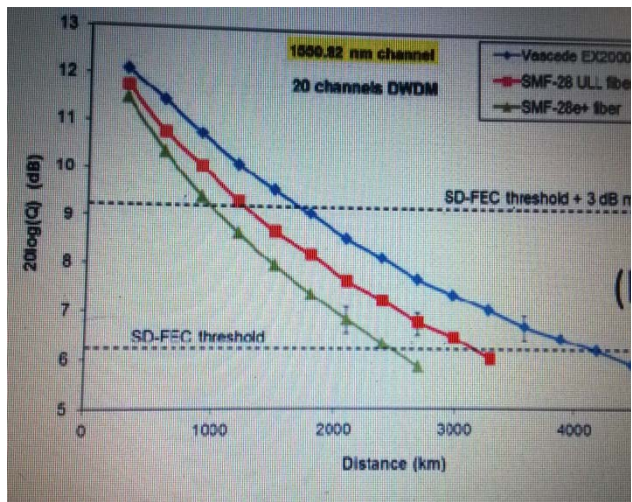


Figure 18(b): Q vs. transmission distance for 1550.92 nm channel in 20 channel systems. Error bars show range of Q values over all 20 channels at select distances.

The Raman enhanced frameworks information are condensed in Figs. 18(a) and 18(b). In Fig. 18(a), unquestionably the achieve lengths of the strands at diverse Q edges are given, while in Fig. 18(b), the achieve lengths are standardized to that of the standard single-mode fiber. with 3 db Q edge, the achieve lengths of the Vascade

Ex2000, SMF-28 ULL, and SMF-28e + fiber frameworks were 1700 km, 1240 km, and 970 km, separately. These results compare to achieve points of interest of the two ultra-low misfortune strands with 3 db edge of 76% and 28%.

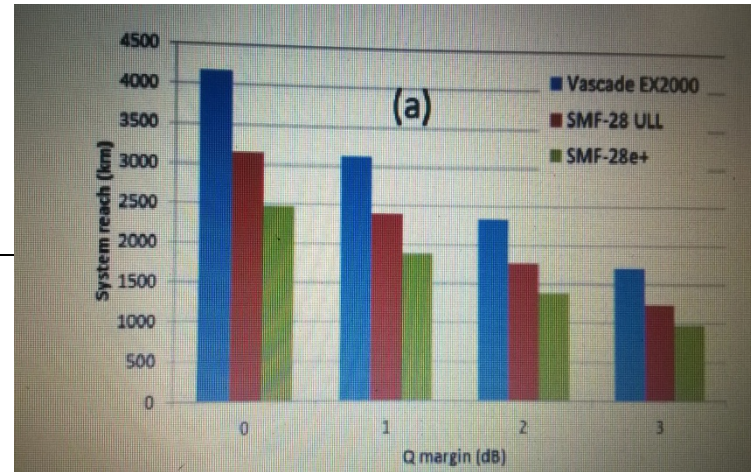


Figure 19(a): Summary of reach length results for Raman systems with 100 km spans. (a) Absolute reach lengths in km.

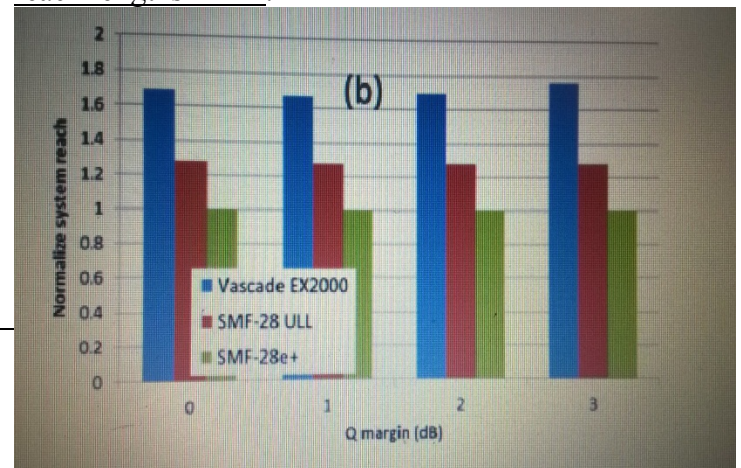


Figure 19(b): Reach lengths normalized to that of standard single-mode fiber.

For these re-flowing circle exploratory frameworks, the reach advantage of the Raman frameworks contrasted with the EDFA frameworks are around half for each of the three optical strands at the 3 db edge

level. It is normal that this achieve playing point would be more noteworthy in straight-line frameworks without extra circle component misfortunes.

#### CONCLUSION:

We have examined 256 Gb/s PM-16qam transmission in excess of three optical filaments with two diverse intensification plans in 100 km compass frameworks. The achieve lengths of every fiber with each one kind of intensification were thought about. For EDFA frameworks, the scope of the standard single-mode fiber was around 630 km with 3 db Q edge over the SD-FEC limit, while a ultra-low misfortune G.652-agreeable fiber span was around 830 km, and it was around 1140 km for a ultra-low misfortune, huge viable region G.654-compliant fiber. For Raman intensified frameworks, those achieve qualities expanded to around 970 km, 1240 km, and 1700 km, separately.

Raman intensification gave around a half reach increase to every one of the three filaments tried. The analyses highlight the profits of utilizing progressed optical filaments with ultra-low weakening and/or bigger compelling range, and also Raman intensification, to attain to achieve lengths with sufficient edge that may be viable and financially savvy in business arrangements.

#### References:

1 G. P. Agrawal, *Lightwave Technology: Telecommunication Systems* (Wiley, Hoboken, NJ, 2005).

2. G. P. Agrawal, *Lightwave Technology: Components and Devices* (Wiley, Hoboken, NJ, 2007).

3. "White Paper"Moving to 100G and Beyond" by Sterling Perrin Senior Analyst, Heavy Reading ,[www.heavyreading.com](http://www.heavyreading.com).

4. "Nonlinear Fiber Optics", G. Agrawal, Academic Press , "Raman Amplifiers for Telecommunications", Edited by M.N. Islam, Springer series in optical sciences, "Raman Amplification for Fiber Communication Systems", Jake Bromage, J. Lightwave Technology Vol. 22 p.79 (2004).

5. I. A. H. Gnauck, P. J. Winzer, S. Chandrasekhar, X. Liu, B. Zhu, and D. W. Peckham, "Spectrally efficient longhaul WDM transmission using 224-Gb/s polarization-multiplexed 16-QAM," *J. Lightwave Technol.* 29(4), 373–377 (2011), M. S. Alfiad, M. Kuschnerov, S. L. Hansen, T. Wuth, D. van den Borne, and H. de Waardt, "Transmission of 11 x 224-Gb/s POLMUX-RZ-16QAM over 1500 km of LongLine and pure-silica SMF," in *Proceedings of European Conf. Opt. Commun.*(2010), paper We.8.C.2, S. Oda, T. Tanimura, Y. Cao, T. Hoshida, Y. Akiyama, H. Nakashima, C. Ohshima, K. Sone, Y. Aoki, M. Yan, Z. Tao, J. C. Rasmussen, Y. Yamamoto, and T. Sasaki, "80x224 Gb/s unrepeated transmission over 240 km of large-Aeff pure silica core fibre without remote optical pre-amplifier," in *Proceedings of European Conf. Opt. Commun.*(2011), paper Th.13.C.7, M. Mussolin, D. Rafique, J. Martensson, M. Forzati, J. K. Fischer, L. Molle, M. Nolle, C. Schubert, and A. D. Ellis, "Polarization multiplexed 224 Gb/s 16QAM transmission employing digital back-propagation," in *Proceedings of European Conf. Opt. Commun.*(2011), paper We.8.B.6, O. Bertran-Pardo, J. Renaudier, H. Mardoyan,

*P. Tan, F. Vacondio, M. Salsi, G. Charlet, S. Bigo, A. Konczykowska, J.-Y. Dupuy, F. Jorge, M. Riet, and J. Godin, "Experimental assessment of transmission reach for uncompensated 32-GBaud PDM-QPSK and PDM-16QAM," in Optical Fiber Communication Conference and Exposition (OFC) and National Fiber Optic Engineers Conference (NFOEC)(Optical Society of America, Washington, DC, 2012), paper JW2A.53, J. Renaudier, O. Bertran-Pardo, H. Mardoyan, P. Tran, G. Charlet, S. Bigo, A. Konczykowska, J.-Y. Dupuy, F. Jorge, M. Riet, and J. Godin, "Spectrally efficient long-haul transmission of 22-Tb/s using 40-Gbaud PDM-16QAM with coherent detection," in Optical Fiber Communication Conference and Exposition (OFC) and*

*National Fiber Optic Engineers Conference (NFOEC)(Optical Society of America, Washington, DC, 2012), paper OW4C.2, K. Roberts, M. O'Sullivan, K.-T. Wu, H. Sun, A. Awadalla, D. J. Krause, and C. Laperle, "Performance of dualpolarization QPSK for optical transport systems," J. Lightwave Technol. 27(16), 3546–3559 (2009), I. Fatadin, D. Ives, and S. J. Savory, "Blind equalization and carrier phase recovery in a 16-QAM optical coherent system," J. Lightwave Technol. 27(15), 3042–3049 (2009).*

# Information and Knowledge Engineering

Okal Christopher Otieno  
Department of Information Technology  
Mount Kenya University  
Nairobi, Kenya

**ABSTRACT: Information and knowledge engineering is a significant field for various applications on processes around the globe. This investigation paper provides an overview of the status of development of the concept and how it relates to other areas such as information technology. The area that is of primary concern to this research is the connection with artificial intelligence. There is a revelation that knowledge engineering derives most of its operational domains from the principles of that concept. There is also a strong relation with the area of software development. As the research shows, they both have the same end products and procedures of attaining it. They both produce a computer program that deals with a particular issue in their contexts. The discussion also focuses on the two modeling approaches that are canonical probabilistic and decision-based software processes. There is a description of the typical knowledge engineering process that each activity has to go through for efficient operation. The paper also takes a look at of the applications of knowledge-based systems in the industry.**

## 1.0 INTRODUCTION

Knowledge-based systems are a form of computer programs that utilize a particular base of information to solve complex problems that their users define to them [12]. They derive their workability from the ideals and principles of artificial intelligence to provide algorithms that develop the architecture a person uses to

describe and solve a problem that is pertinent to a particular field of study. Therefore, the utilization of knowledge-based systems can lie in the technical, social and scientific fields of approach depending on how the developer creates them. There have been various adjustments on these systems to ensure they advance according to the other technological signs of progress that scientists are making in computing [1]. There are various attempts, for example, to make the knowledge-based systems perform their function while using the internet. Such an improvement can make them relevant to cloud computing platforms whose essence in the fields of information technology has been steadily growing in the recent past [17].

This paper has its focus on the techniques that the people use to create and manage knowledge-based systems. Knowledge engineering covers the technical, social and scientific fields to develop, maintain and utilize those systems and their pertinent aspects [19]. There are various enhancements that the systems bring to the area of information technology that leads to the increase in their application. Knowledge engineering derives from different aspects of that field including databases, artificial intelligence, expert systems, and decision support systems [11]. The concept is important in shared frameworks of information that enable its broad utilization. This paper discusses the concepts of information and knowledge engineering and its relationship with computing. There will be an overview of the functions of knowledge-based systems and

their functions in the field of information technology and management to establish these concepts. Then there will also be a look at the models that define knowledge engineering.

## 2.0 LITERATURE REVIEW

### 2.1 *What is Information and Knowledge Management?*

Knowledge engineering is a multidisciplinary concept that requires the combination of the social, technical and scientific aspects to develop, utilize and maintain a knowledge-based system [22]. The concept applies information from the development of computing systems as a basis for its existence. Knowledge engineering, therefore, has a close relationship with most other fields of computer science that this paper will focus upon in the next sections. For example, the approaches that knowledge engineering uses derive from the principles of artificial intelligence that is a developing concept in the field of computing [14]. In other words, knowledge engineering is a field that leads to the development of computer programs and systems that facilitate problem solution using artificial intelligence. Therefore, it also derives from the basis of software engineering to develop the end product that consumers wish to see. It is then easy for the researcher to conclude that the concept is a sub-domain of computing and information technology.

There has been an increase in societal dependence on information systems and knowledge as a way to develop various models that assist a lot of people's operations [21]. Those actions have necessitated the growth in the fields of knowledge engineering to provide solutions to the workability of those systems and their aspects in different areas of study. That has led to the evolution of the knowledge-based systems and the field of knowledge

engineering to adapt to the changes that the society is witnessing. Technological advancements have been at the forefront of the factors that have the greatest influence on the development of knowledge engineering [4]. The significance of knowledge in businesses and other professions have also had another influence as it enhances the efficiency of processes in those areas. The dynamism of knowledge engineering in covering most fields from science to sociology place it at the center stage as the solution provider to the people in those practices. An example of the evolution is that the knowledge-based systems now have to be able to use the internet as a platform for operation. That is a break from the traditional approaches that relied on a local network alone; therefore, technology has an impact. Most people are tending towards the use of cloud computing due to the benefits it offers that are better than the conventional models [3]. That proposition also influences an evolution of knowledge engineering to make it relevant to the issues that people wish to solve.

### 2.2 *Relationship with Artificial Intelligence*

There is a relationship between the concept of and knowledge engineering and artificial intelligence AI. It is important to outline the definition of AI to determine which of the two concepts influences the other to establish this connection. Therefore, artificial intelligence refers to the ability of a machine and its software and hardware components to perform logic and abstract functions including communication between various systems [13]. When dealing with the information and the systems, the definition of that concept becomes the capacity to perceive and retain that knowledge to facilitate problem-solving strategies [6]. Most other description of artificial intelligence reflects it as a mental ability in the machines that enables them to reason



through a problem, plan on the pertinent activities and provides solutions through abstract thought. There are big variations from the human intelligence that arise from the better abilities that the people have to adapt their conscious to different stimuli. A man can change their thoughts even when the factors shift entirely and make decisions that fit into that context. On the other hand, artificial intelligence only deals with the variables that the machine has its inputs to adjust within a particular context [26]. For example, a robot that assembles vehicles has better accuracy than a human but, it cannot run when a fire occurs in the building because that function is not in its algorithm. Therefore, the beings of artificial intelligence can only vary their conscious within the constraints that the developers input into their operations.

As per the definitions of knowledge engineering above, it is a concept in computing that develops systems that solve problems using particular information and algorithms. Most scholars record that the whole field of knowledge engineering derives from earlier developments, expert systems, that relied on artificial intelligence [10]. The expert systems were an attempt by computer scientists to simulate the intelligent capabilities in humans and enable machines to substitute the people-effort in problem-solving exercises [7]. The strategy aimed at providing a computing system with adequate information through the input of algorithms that define certain functions. Afterward, the user provides some problems for the machine to determine how it is capable of utilizing that knowledge and providing a solution to it. Therefore, a knowledge-based system comprises of two important parts that enable its efficient functioning. The first is the knowledge base that contains all the information and algorithms and information that define the procedures it is likely to use to solve a

problem [9]. This part contains all the facts and formulae that the system will select from according to the input that it will get about a particular issue from the user. It is similar to the way that a human being retains the information they get from books and other sources and turn it into the knowledge that they use to deal with their issues. The second part of the knowledge-based system is the reasoning or inference processor that enables the execution of the information and procedures in the knowledge base. They are a reactive process to the input that the system recognizes as a problem, and they initiate the whole program to deal with the functions that the user brings to it.

The way that the knowledge-based system functions mirrors the exact conceptualization that defines artificial intelligence. The system requires an input about the procedures that they will follow to execute the information that they have to complete a certain problem-solving task [5]. They follow certain procedures that the developer predefines to allow the system to provide a solution that is pertinent to the input they receive from the end-users. Therefore, they rely on the principles of artificial intelligence for operation with the inability to adjust their activities outside the scope of the knowledge base. If, for example, a person tampers with the knowledge case, the systems do not have the ability to understand that act and restore their operations to the initial convention. They will just follow the new algorithm without knowing if it is harmful or not. There have been broad developments in the field of artificial intelligence through the involvement of scientists in the research. There have also been significant improvements that imply enhancements in knowledge engineering. As artificial intelligence keeps growing in its capacity, so are the capabilities of knowledge-based systems. There is an always increasing



dependence on the knowledge that is available to facilitate further development in different areas of study. That is one of the factors that motivate the rise in efforts to develop better systems that can provide sufficient strategies to deal with human issues.

The objectives of knowledge and software engineering are similar as they pertain to the same context and also present an equal end product for the users. They all hope to develop a computer program that a user can apply in developing solutions to their problems. They both are a development from the traditional computing by bringing a conceptual approach on board that utilizes available knowledge to produce more.

### *2.3 The Functioning of Knowledge Engineering and Process*

The understanding of how knowledge-based systems function provides an excellent opportunity to get insights on how the concepts of knowledge engineering work. This section presents an overview of the typical process that the exercises in this field follow to arrive at a strategic solution to the problem a user gives. The first step in the functionality of the systems is to identify the type of tasks that the user inputs and needs a solution [21]. In this case, the developer defines the type of questions that the system will answer and sets certain rules on checks that it should perform on the user's input. This coding enables the knowledge base to determine if it has the relevant tools that can solve that problem before it proceeds to the next step.

The second step in the functionality of a knowledge engineering process is to gather the information that it has in its base to determine its pertinence to the rules of problem-solving [25]. The system reflects the standards that the developer defined in the knowledge base about a particular problem to identify the correct strategy and

approach to the problem. The system uses this step to determine if there are more sets of rules that it needs to provide a better solution or the current ones are sufficient. It is in this phase that it might ask for further assistance to clarify the problem or provide further variables from the humans to facilitate the processing. Take an example of a system that performs calculations, this stage helps the system to detect that the user has input letters instead of numbers. It then informs the user to enter figures or define an algorithm that calculates with letters.

The third step in the knowledge engineering process function is the development of ontologies that are a set of vocabularies that describes the objects that the user may be interested in for the problem at hand [16]. They include constants, variables, predicates, and functions that the program will use to deal with the issues that it identifies from the steps above. Ontology allows the different agents of the problem solution to share the information that is available in the knowledge base to determine the best strategy for solving the problem. They also allow humans to give a hand to ensuring the system uses the best path towards the solution. For example, an ontology will all the program to know that it should use pi to find the area of a circle instead of sigma as both may exist in the knowledge base. It then tells the program on the further steps it should conduct, say multiplication instead of addition.

Encoding the knowledge from the other phases represents the fourth stage of a process in knowledge engineering [2]. This step involves the identification of the axioms that are pertinent to the domain of the problem that the user wishes to solve. The system tries to establish the relationship between the input that the user provides, and they type of output that they expect to get after a computation. Therefore, this phase involves the development of certain

algorithms that can make up a particular formula that provides the solutions that the end-users anticipate. It is at this stage that the program should identify any mistakes that it could make in the phase of developing the ontologies in the previous stages. It provides a correction on those issues and then creates the axioms of the problem domain. In a proper perspective, this is the stage at which a human could differentiate between the formula for calculating the area of a circle and that for its circumference. They may use the same parameters for measurement, but the method isolates the order of application to identify the correct solution. It gauges between the area and the circumference to derive the right information from the knowledge base.

The fifth step in this process involves the encoding of the instances that will lead to a solution of the problem [20]. From the other steps of the activity, the program has all the information that it needs to solve the problem. But, it will go back to the inputs once more to determine what values fit in what context. It is at this stage that it outlines how the combination of axioms from the previous stage matches with the data that the user inputs to develop a solution that is pertinent to their problem. Going back to the example of the circle, this is where the program differentiates the values into those that represent the radius and diameter. This phase enables it to decide the best procedures for ensuring the formula that it chooses adhere with the right variables to create the relevant solution. By the end of the activities of this stage, the program should be in a position to provide a basic solution to some problems. After the completion of this stage, the developers move on to testing the application to ensure it is safe for use.

The sixth step involves the activities of finalizing with the program [24]. It has been capable of following the procedures

that will solve the problems in its fields and requires a confirmation of its efficiency. Therefore, the developers present a set of challenges to the knowledge base to ensure it has all the correct information and procedures that will deal with the issues. They create the instances that can relate to the typical problems that are likely to occur in the program and debug it. There has to be a focus on the possible complexities that any user can bring on to ensure that the knowledge-based system is capable of sustaining various amounts of operational pressure. The result of this phase should provide a program that has all the efficiencies that the end users define in their requirements. It is at this procedural stage that the developer gets the opportunity to detect and eliminate all errors that can minimize the optimal performance of the knowledge-based system.

#### *2.4 Models for Knowledge Engineering*

There are various approaches that the developers of a knowledge based system may wish to use to outline the objectives of the process above. The modeling strategies are necessary for different contexts to ensure that the product is pertinent to the needs of the target users. This paper focuses on the canonical Probabilistic and also the Decision-Based Software Process Models.

##### *2.4.1 Canonical Probabilistic Model of Knowledge Engineering*

Probabilistic models provide the best avenue that most analysts can use to deal with large data sets that are tough to predict the outcomes. They use artificial intelligence as a basis to develop the techniques of analysis, and this approach increases their efficiency. Canonical probabilistic approaches to knowledge engineering provide an opportunity for the developers to predict and satisfy the demands that numerous and very random people are likely

to create. Therefore, the approach focuses on establishing the possible outcomes and responses towards a particular stimulus and uses that information to create the knowledge-based systems. The challenge in the application of this type of modeling in knowledge engineering has been to get the numerical data that can provide sufficient information for the development [8]. The canonical probabilistic modeling uses various techniques and approaches that rely heavily on mathematical formulae for success. Therefore, there has to be sufficient quantitative data about the target users if the developers are to get accurate and positive outcomes from their activities. Sometimes people use case studies of particular applications to provide a ground for the anticipation of the results before indulging in the development.

#### *2.4.2 Decision-Based Software Process Model of Knowledge Engineering*

This approach uses the participation of all the stakeholders that are likely to use the program to provide the information about their requirements [23]. Then, the developers will use that information to decide on the best approaches that can help the people to solve their problem. The modeling of the knowledge-based system derives from that decision that may even break further down to provide more specific solutions. The model requires that the developers have to be in constant consultation with representatives of the stakeholders to determine how each of the operations has to proceed. It is the only way to ensure that the final product is pertinent to the requirements. The approach focuses on the organization elements that include the people, their tasks, their operational structure, information technology, and the techniques of production. The decision-oriented modeling eliminates the possibility

of marginal errors in the process that can be costly for the clients and the developers.

#### *2.5 Common Applications of Knowledge-Based Systems*

One of the areas that widely use knowledge-based systems is the scheduling of processes in activities such as manufacturing and other industrial processes [18]. Such activities require the optimization of resources, time, and costs to ensure the actor achieve the best profits from their operations. The difficulties come in the instances that involve exponential constraints and multiple variables that need simultaneous computations and solutions. But the process has prior information about what combination of activities failed and those that are a success. Then they use that information to develop a system that can provide an optimization strategy to handle the issues. The system requires the ability to change the schedule as the requirements shift [15]. That exploration is one important application of knowledge-based systems. The basics of those applications also extend to other areas of industry including information technology. Therefore, knowledge-based systems are an important contributor to various activities in business and other fields.

#### *Conclusion*

The concept of knowledge and information engineering is an essential contributor to the development of the world that is increasing its dependency on computing systems. The study sought to establish the functions of knowledge engineering and indulged into the canonical probabilistic and decision-based models that developers can use to create knowledge-based systems. There is a good relationship between knowledge engineering and artificial intelligence as the latter provides the basis upon which knowledge-based systems develop. Therefore, it is easier to

conclude that any progress that researchers and scientists make in the field of computing will also influence the development of knowledge engineering. That is because information and knowledge engineering also derives from other concepts such as software development that are realizing rapid growth in the current society. The applications of knowledge-based systems are also essential to most functions in industry and businesses. People would be using the programs without even realizing they are engaging in the concepts of information and knowledge engineering. The applications are significant to the extent of influencing the procession activities in areas such as manufacturing and other industries such as mining. Therefore, knowledge engineering is an important concept that requires more attention and development to benefit the society.

## References

- [1].Akerkar, Rajendra, and Priti Sajja. *Knowledge-based systems*. Jones & Bartlett Publishers, 2010.
- [2].Baumeister, Joachim, Jochen Reutelshoefer, and Frank Puppe. "KnowWE: a Semantic Wiki for knowledge engineering." *Applied Intelligence* 35.3 (2011): 323-344.
- [3].Beloglazov, Anton, Jemal Abawajy, and Rajkumar Buyya. "Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing." *Future generation computer systems* 28.5 (2012): 755-768.
- [4].Brodie, Michael L., and John Mylopoulos, eds. *On knowledge base management systems: integrating artificial intelligence and database technologies*. Springer Science & Business Media, 2012.
- [5].Cheung, Chi Fai, C. M. Cheung, and S. K. Kwok. "A knowledge-based customization system for supply chain integration." *Expert Systems with Applications* 39.4 (2012): 3906-3924.
- [6].Cohen, Paul R., and Edward A. Feigenbaum, eds. *The handbook of artificial intelligence*. Vol. 3. Butterworth-Heinemann, 2014.
- [7].David, Jean-Marc, Jean-Paul Krivine, and Reid Simmons, eds. *Second generation expert systems*. Springer Science & Business Media, 2012.
- [8].Diez, F. Javier, and Marek J. Druzdzel. *Canonical probabilistic models for knowledge engineering*. Technical Report CISIAD-06-01, UNED, Madrid, Spain, 2006.
- [9].Eldrandaly, Khalid, and Soad Naguib. "A knowledge-based system for GIS software selection." *Int. Arab J. Inf. Technol.* 10.2 (2013): 152-159.
- [10]. Hwang, Gwo-Jen, et al. "A knowledge engineering approach to developing educational computer games for improving students' differentiating knowledge." *British Journal of Educational Technology* 44.2 (2013): 183-196.
- [11]. Kasabov, Nikola K. *Foundations of neural networks, fuzzy systems, and knowledge engineering*. Marcel Alencar, 1996.
- [12]. Li, B. M., S. Q. Xie, and X. Xu. "Recent development of knowledge-based systems, methods and tools for One-of-a-Kind Production." *Knowledge-Based Systems* 24.7 (2011): 1108-1119.
- [13]. Nilsson, Nils J. *Principles of artificial intelligence*. Morgan Kaufmann, 2014.
- [14]. Pham, Duc Truong, ed. *Artificial intelligence in design*. Springer Science & Business Media, 2012.
- [15]. Raman, RV, and K. V. K. Prasad. "Applications of Knowledge Based Systems in Mining Engineering." *APCOM 87: Mining* 1. 1987.

- [16]. Rao, Lila, Gunjan Mansingh, and Kweku-Muata Osei-Bryson. "Building ontology based knowledge maps to assist business process re-engineering." *Decision Support Systems* 52.3 (2012): 577-589.
- [17]. Sarathy, Vijay, Purnendu Narayan, and Rao Mikkilineni. "Next generation cloud computing architecture: Enabling real-time dynamism for shared distributed physical infrastructure." *Enabling Technologies: Infrastructures for Collaborative Enterprises (WETICE), 2010 19th IEEE International Workshop on*. IEEE, 2010.
- [18]. Saucer, J. "Knowledge-based systems techniques and applications in scheduling." *Knowledge-Based Systems Techniques and Applications, TL Leondes, ed., San Diego, Academic Press*. 1999.
- [19]. Schreiber, Guus. *Knowledge engineering and management: the CommonKADS methodology*. MIT press, 2000.
- [20]. Simpson, Ron M., et al. "GIPO: an integrated graphical tool to support knowledge engineering in AI planning." *Sixth European Conference on Planning*. 2014.
- [21]. Sriram, Ram D. *Intelligent systems for engineering: a knowledge-based approach*. Springer Science & Business Media, 2012.
- [22]. Studer, Rudi, V. Richard Benjamins, and Dieter Fensel. "Knowledge engineering: principles and methods." *Data & knowledge engineering* 25.1 (1998): 161-197.
- [23]. Toffolon, Claudine, and Salem Ben Dhaou Dakhli. "A Decision-oriented model of software engineering processes." (2007).
- [24]. Tzafestas, Spyros G., ed. *Knowledge-based system diagnosis, supervision, and control*. Springer Science & Business Media, 2013.
- [25]. Warner, Homer R., Dean K. Sorenson, and Omar Bouhaddou. *Knowledge engineering in health informatics*. Springer Science & Business Media, 2012.
- [26]. Wenger, Etienne. *Artificial intelligence and tutoring systems: computational and cognitive approaches to the communication of knowledge*. Morgan Kaufmann, 2014.

# Online Support vector machines based on the data density

Saeideh beygbabaei

Department of computer

Zanjan Branch, Islamic Azad University

Zanjan, Iran

**Abstract**— nowadays we are faced with an infinite data sets, such as bank card transactions, which according to its specific approach, the traditional classification methods cannot be used for them. In this data, the classification model must be created with a limited number of data and then with the receiving every new data, first, it has been classified and ultimately according to the actual label (which obtained with a delay) improve classification model. This problem known the online classification data. one of the effective ways to solve this problem, the methods are based on support vector machines that can pointed to OISVM, ROSVM, LASVM and... .in this classification accuracy and speed and memory is very important; on the other hand, since finishing operations support vector machines only depends to support vector which is nearly to optimal hyperplane clastic; all other samples are irrelevant about this operation of the decision or optimal hyperplane, in which case it is possible classification accuracy be low. In this paper to achieve the desirable and accuracy and speed memory, we want by reflect the distribution density samples and linearly independent vectors, improve the support vector machines. Performance of the proposed method on the 10 dataset from UCI database and KEELS evaluation.

**Keywords:** support vector machines, linear independent vector, relative density degree, online learning

## I. INTRODUCTION

Support vector machines (SVMs) [1] are one of the most reputable and promising classification algorithms. SVM foretaste that one sample Classified in which class or group. The algorithm for the separation of groups from each use the page. As opposed to another learning methods such as, e.g., neural networks, they are forcefully theoretically founded, and have been shown to enjoy excellent performance in various applications [2]. one framework, though, in which their power has not yet been fully developed is on-line learning. Online Training is determined as an algorithm allows several incremental updates of a model to be processed .SVM Assuming that the data are available and once learns data and applies to predict the other data. But online data are regularly are producing and may the data changed over time. Some data newly are produced and this is defect SVM. Our problem is how can change SVM To be able with use the data density, learn the data as online.

## II. TOPICS STUDIED IN THIS PAPER

### A. online Classifiers challenges

Classification problems used in different areas such as data analysis, machine learning, and data mining and statistical inference, classification Methods are in result a form of supervised learning methods which a set of dependent variables must be estimated based on the input feature set. Many proposed methods for the classification assume that the data sets are static and to create models of the data can be used even have perform multiple passes over on data. However, having multiple passes on the data to create a model of data mining in the online data is not possible. Moreover, one of the most momentous challenges in the classification online data, is discussion of not having previous knowledge an online data that causes commonly cannot be used older methods for online data. The most important challenge in the classification online data discussion is the accuracy and speed and memory in data. We have to resolve the challenges Training Speed And savings in memory and Accuracy suggest algorithm Online support vector machine based on data density, The algorithm calculates density of sample each sample entering and applies a set of linearly independent observation and also we use Newton to update the method, That reduces the time and space required. This method also has good speed and accuracy.

### B. support vector machines

This method is among a relatively new methods that have advantage performance in recent years compared to older methods for classification Based Working Classification SVM, classify linear data. We are trying in partition the linear data select the linear that have more confident marginally. Solve the equation to find the optimal line for data are done by QP methods well-known methods in solving restricted problems. SVM is a binary classifier that separates two classes by using a linear border. In this way, using every bands and an optimization algorithm, to obtain samples that form the boundaries of classes, these samples called support vectors. Number of training places that are nearest to the decision boundary can considered as a subset to define the boundaries of decision and support vectors. We have Training data set D includes n elements that can be defined as follows:

$$D = \{(x_i, y_i) | x_i \in R^p, y_i \in \{-1, 1\}\}_{i=1}^n \quad (1)$$

Consider samples can be separate in  $R^m$  by a linear function. Such  $f(x) = w \cdot x + b$ ,  $w \in R^m$ ,  $b \in R$  SVM algorithm finds a function, which is  $\|w\|$  minimum. Provided that  $y_i(w \cdot x_i + b) - 1 \geq 0$ . If the samples are not possible to linearly separated, L slack variables  $\xi \geq 0$  are introduced and

$$\operatorname{argmin} \frac{1}{2} \|w\|^2 + c \sum_{i=1}^l \xi_i^p \quad (2)$$

is sought for, subject to the limitation  $y_i(w \cdot x_i + b) \geq 1 - \xi_i$ , where  $c \in R^+$  is an error penalty coefficient and P is usually 1 or 2. The problem is compactly expressed in Lagrangian form by further introducing L pairs of coefficients  $\alpha_i, \mu_i$  and then minimizing

$$l_p = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i (y_i(w \cdot x_i + b) - 1 + \xi_i) + c \sum_{i=1}^l \xi_i^p - \sum_{i=1}^l \mu_i \xi_i \quad (3)$$

Subject to  $\alpha_i, \mu_i \geq 0$  Using the Karush–Kuhn–Tucker (KKT) optimality conditions, we obtain that

$$\frac{\partial l_p}{\partial w} = w - \sum_{i=1}^l \alpha_i y_i x_i = 0 \quad (4)$$

That is

$$w = \sum_{i=1}^l \alpha_i y_i x_i \quad (5)$$

Hence the approximating function  $f(x)$  can be expressed as

$$f(x) = \sum_{i=1}^l \alpha_i y_i x \cdot x_i + b \quad (6)$$

To improve the discriminative power of an SVM, The  $x_i$ 's are generally mapped to a high, possibly infinite dimensional space (the feature space) via an on-linear mapping  $\Phi(X)$ ; the core of the SVM becomes then the so called kernel function  $k(a, b) = \Phi(a) \cdot \Phi(b)$ . The kernel matrix K is defined beside such that  $k_{ij} = k(x_i, x_j)$  In fact, dimension the core is the dimension characteristics space. Widely used kernels are the polynomial one (finite- dimensional) and the Gaussian one (infinite-dimensional). In the end, Eq. (6) is rewritten as  $f(x) = \sum_{i=1}^l \alpha_i y_i k(x, x_i) + b$  After minimization of  $L_p$ , some of the  $\alpha_i$ 's (really most of them in many applied applications) are equal to zero; those  $x_i$ 's for which this does not hold are called support vectors. The solution depends on them alone and their number is proportional to the number of training samples [3]. The standard SVM algorithm is meant to be used batch-wise; to increase it to the on-line setting two different approaches have been Offered: (i) the batch algorithm is adapted to test one sample at the time and produce a new approximate solution; (ii) accurate methods that incrementally update the solution. In both cases we have that the potentially endless flow of training samples of the on-line setting will bring sooner or later to an explosion of the number of support vectors, and hence of the testing time.

### C. on-line independent SVMs

In SVMs When we have a never-ending stream of data because of space and time are not suitable requisition for online learning. In standard learning, a set of (sample, label) pairs drawn from an unknown probability distribution is given in advance (training set); the work is to find a function (hypothesis) such that its sign supreme determines the label of any future sample drawn from the same distribution. As

opposed to this, in on-line learning samples and labels are made available in time, so that no knowledge of the training set can be supposed a priori. The hypothesis must therefore be built incrementally every time a new sample is available. Let us call this operation of building a new hypothesis, a round. Formally, let  $\{x_i, y_i\}_{i=1}^l$ , with  $x_i \in R^m$ ,  $l \in R^+$  and  $y_i \in \{-1, 1\}$ ; be the full training set, and let  $h_i$  denote the hypothesis built at round i, when only the (sample, label) pairs up to i are available. At the next round, a new sample  $x_{i+1}$  is available and its label is predicted using  $h_i$ . The true label  $y_{i+1}$  is then matched against this prediction, and a new hypothesis  $h_{i+1}$  is built taking into account the loss incurred in the prediction. In the end, for any given sequence of samples  $(x_1, y_1), \dots, (x_l, y_l)$ , a sequence of hypotheses  $h_1, \dots, h_l$  is built such that  $h_i$  depends only on  $h_{i-1}$  and  $(x_i, y_i)$ . Note that any standard machine learning algorithm can be adapted to work in the on-line setting only retraining from scratch each time a new sample is acquired. However, this would result in an extremely inefficient algorithm. In the following we sketch the theory of SVM that gives us the tools to extend it to the on-line setting in an efficient way.

#### 1) density of sample

The relative density degree for a sample Indicative how dense the region in which the corresponding sample locates is compared to other region in a given data set. Here we focus on assigning each point a relative margin. The relative margins of points want to be optimized by algorithms, and have no relation with the density distribution According to the density distribution of a given data set, a data point in it may locate in a dense region and has a higher density degree, or in a Non condensing region and has a lower density degree. The final decision function of SVMs just depends on support vectors which lie closet to the optimal separating hyperplane, whereas all other samples are irrelevant to this decision function. If the given data are smoothness, or satisfy the low density separation assumption, resulting SVs usually locate in a lower density region. However, samples with higher density degrees should be included in the representation of decision function in order to more correctly classify the given data set [4]. For the unsmooth data, resulting SVs may locate in a lower density region or not. But the ‘‘optimal’’ separating hyperplane just based on SVs, without considering the density distribution, may not be the optimal actually. Therefore, we want to reflect the density distribution of data in SVMs. We extract relative density degrees of training data as the density information and assign them to the corresponding data points as relative margins. we use our proposed method to extract relative density of the samples of K-nearest neighbor method such as in [5]. In this way, By entering Each sample we calculated density the following way :  $X_m = \{x_{mj}\}_{j=1}^{lm}$ ,  $m = 1, 2, \dots, c$  and the value of K, we search K nearest neighbors for  $x_{mj}$ , in the m-th class by using some distance metric  $d(x_{mj}, x_m)$ . Let  $x_{mj}^k$  be the K-th nearest



neighbor,  $d(x_{mj}, x_{mj}^k)$  be the distance between  $x_{mj}$  and  $x_{mj}^k$ , and

$$D_m^k = 1/lm \sum_{j=1}^{lm} d(x_{mj}, x_{mj}^k) \quad (7)$$

The relative density degree  $\rho_{mj}$  for  $x_{mj}$  is defined by

$$\rho_{mj} = \frac{D_m^k}{d(x_{mj}, x_{mj}^k)} \quad (8)$$

## 2) Exploiting linear independence sample

A system involved in on-line learning must face a potentially endless flow of training data, updating its knowledge after every new sample. This setting is particularly hard for SVMs as the size of an SVM solution grows linearly with the number of training samples taken into account. Since any real system has access to finite resources (e.g., finite computational power, memory, etc.), a strategy to limit the number of data points is needed, and a trade-off to accuracy must be accepted. So, it becomes decisive to find a way to save resources while obtaining an acceptable approximation of the ideal system. Consider  $f(x) = \sum_{i=1}^l \alpha_i y_i k(x, x_i) + b$ . The representation of  $f(x)$  is created by summing up as many factors as support vectors. Really, one step further can be taken: if some of the support vectors are linearly dependent on the others in the feature space, some of them can be expressed as a function of the others, so reducing the expression of  $f(x)$ . Denote the indices of the vectors in the current basis, after  $L$  training examples, by  $B$ . When the algorithm receives  $x_{l+1}$  it has to check if it is linearly independent or not from the basis vectors. Generally, checking whether a matrix has full rank is done via some decomposition, or by looking at the eigenvalues of the matrix; but here we want to check whether a single vector is linearly independent from a matrix of vectors already known to be full rank. It is then simpler to exploit the definition of linear independence and check how well the vector can be approximated by a linear combination of the vectors in the matrix. let  $d_j \in \mathbb{R}$ ; then let

$$\Delta = \min_d \left\| \sum_{j \in B} d_j \phi(x_j) - \phi(x_{l+1}) \right\|^2 < \eta \quad (9)$$

If in the formula  $\Delta > 0$ ,  $x_{l+1}$  is linearly independent with respect to the basis, and  $L+1$  is added to  $B$ .

In practice, one must check whether  $\Delta < \eta$  where  $0 < \eta$  is a tolerance factor, and expect that larger values of  $\eta$  lead to worse accuracy, but also to smaller bases. If  $\eta$  is set to zero the solution found will be the same as in the classical SVM formulation; therefore, no approximation what so ever is involved, unless one gives it up in order to get even fewer support vectors. An impressive way to evaluate  $\Delta$  is needed.

to expand (9) and remembering the compliment of the kernel matrix  $K$ , we get

$$\Delta = \min_d (d^T k_{BB} d - 2d^T k + k(x_{l+1}, x_{l+1})) \quad (10)$$

Applying the extremum conditions with respect to  $d$  to (10) we obtain that  $d^* = k_{BB}^{-1} k$  and, by replacing this in (10) once,

$$\Delta = k(x_{l+1}, x_{l+1}) - k^T d^* < \eta \quad (11)$$

Note that  $K_{BB}$  can be safely inverted since, by incremental construction, it is full rank. An efficient way to do it, exploiting the incremental nature of the approach, is that of updating it recursively. Using the matrix inversion lemma, after the addition of a new sample the new  $k_{BB}^{-1}$  becomes

$$\begin{bmatrix} k_{BB}^{-1} & \vdots \\ \dots & \dots \end{bmatrix} + \frac{1}{\Delta} \begin{bmatrix} d^* \\ -1 \end{bmatrix} [d^{*T} \quad -1] \quad (12)$$

Therefore, with a set of independent training vectors of training samples as the basis vectors and using a Hilbert space and theorem introduced the support vector machine. In this way becomes:  $\arg \min_{w,b} \frac{1}{2} \|w\|^2 + c \sum_{i=1}^l \xi_i^p$ ,  $w$  can be write  $w = \sum_{i=1}^l \mathcal{B}_i \phi(x_i)$ . The expression kernel matrix and norm two  $W$ . The formula for change such as in [6]:

$$\|w\|^2 = \sum_{i,j=1}^l \mathcal{B}_i \mathcal{B}_j \phi(x_i) \phi(x_j) = \sum_{i,j=1}^l \mathcal{B}_i \mathcal{B}_j K_{ij}$$

**The main problem with the use of density that uses algorithm KNN Achieved; change as follows.**

$$\arg \min_{\mathcal{B}, b} \frac{1}{2} \sum_{i,j=1}^l \mathcal{B}_i \mathcal{B}_j K_{ij} + c \sum_{i=1}^l \xi_i^p \quad (13)$$

$$\text{subject to } y_i \left( \sum_{j=1}^l \mathcal{B}_j K_{ij} + b \right) \geq \rho_i - \xi_i \quad \forall i=1, \dots, l$$

In this section we interfaces for optimization problem (13) without expressing its dual Lagrange and formulas express. Instead we use a number of coefficients of selected basis vectors. So we need a way to optimize the initial formula of the equation (13). The method is selected, consistent by the Newton. If  $D \subset \{1, \dots, l\}$ ,

$$\arg \min \frac{1}{2} \mathcal{B}^T k_{DD} \mathcal{B} + \frac{1}{2} c \sum_{i=1}^l \max(0, \rho_i - y_i k_{iD} \mathcal{B})^2 \quad (14)$$

We then set  $D = B$ , which assures that the solution to the problem is unique, since  $K_{BB}$  is full rank by construction. When a new sample  $x_{L+1}$  is available, the Newton method goes as follows:

1. use the current value of  $B$  as starting vector;
2.  $o_{l+1} = k_{l+1, B} \mathcal{B}$  if  $\rho_i - y_{l+1} o_{l+1} \geq 0$  then stop: the current solution is already optimal. Otherwise,
3. if  $I = \{i: \rho_i - y_i o_i > 0\}$  where  $o_i = k_{i, B} \mathcal{B}$  is the output of the  $i$ -th training sample;
4. update  $B$  with a Newton step:  $\mathcal{B} - \gamma P^{-1} g \rightarrow \mathcal{B}$  that

$$P = k_{BB} + ck_{BI}k_{BI}^T$$

$$g = k_{BB}B - ck_{BI}(y_I - o_I)$$

5. let  $I^{new} = \{i: \rho_i - y_i o_i > 0\}$ , where  $o_i$  are recalculated using new B. if  $I^{new}$  is equal to I stop; otherwise  $I^{new} \rightarrow I$ , and go to step 4.

Generally the algorithm is summarized as follows:

- A. Each sample enters its density is calculated.
- B. Check the samples are linearly independent or not, if the sample is linearly independent add to basis vectors.
- C. Optimization is done gradually.

Algorithm1. Pseudo-code of our method

```

Parameters:  $\eta$ 
Initialization:  $B = \{\}$ 
For each time step  $t=1 \dots L$ ,  $m=1, \dots, c$  do
     $d(x_{mt}, x_{mt}^k)$ 
     $D_m^k = \frac{1}{lm} d(x_{mt}, x_{mt}^k)$ 
     $\rho_i = \frac{D_m^k}{d(x_{mt}, x_{mt}^k)}$ 
     $K = k(x_i, x_j)$ 
     $d = k_{BB}^{-1}k$ 
     $\Delta = k(x_t, x_t) - k^T d$ 
    If  $\Delta \geq \eta$  then
         $B = [B, t]$ 
    End if
     $o_t = k^T B$ 
    If  $o_t \leq \rho_i$ 
         $I^{new} = \{i: \rho_i - y_i o_i > 0\}$ 
        Repeat
             $I = I^{new}$ 
             $P = K_{BB} + cK_{BI}K_{BI}^T$ 
             $B = cP^{-1}K_{BI}y_I$ 
            Recalculate  $o_i, i=1, \dots, t$ 
             $I^{new} = \{i: \rho_i - y_i o_i > 0\}$ 
        Until  $I = I^{new}$ 
    End if
End for
    
```

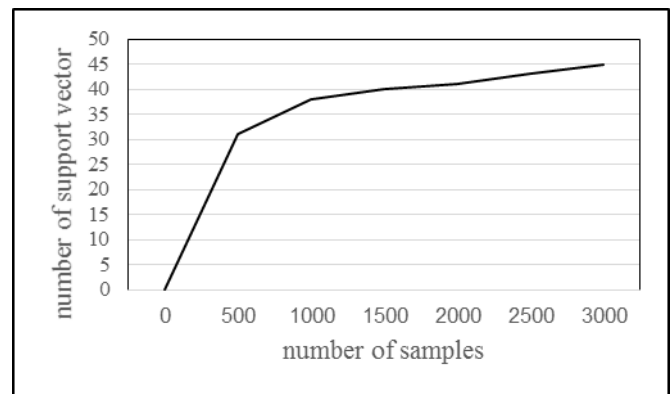
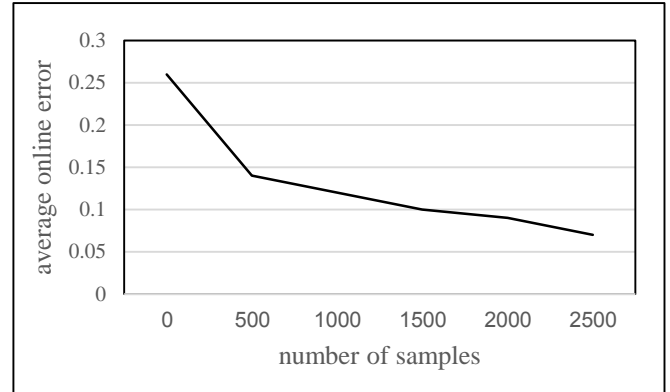
#### D. Experimental results

Results obtained on the average 10 times random performance of data. And in all cases considered the value for K, 10 neighbors. And accuracy is  $\eta = .01$  also kernel used in all performances is Gaussian.

TABLE 1: RESULTS OF TESTS ON THE 10 DATASETS FROM KEELS DATABASE

database	accuracy	error	Numb support vector
Banana	90.716	9.2840	46
Breast	77.72	22.28	41
Diabetes	79	21	11
Flare	69.35	30.65	15
German	78.98	21.02	55
Heart	85.99	14.01	16
Ring norm	98.3	1.7	100
Titanic	78.55	21.5	13
Two norm	97.6	2.40	67
Waveform	90.2	9.80	81

The below images shows the comparison Growth of support vectors and error rate on the banana database With two features and 2500 samples a random selection from 5300 sample. According to the image we find that the error rate is reduced by the time.



#### E. Comparison

Ultimately we show the following table, compared results of tests on the 10 data sets from the UCI and keels database. we have shown Average accuracy of classification Also, the number of support vectors in parentheses. And we compared our method With LIBSVM, OISVM and Incrsvm. We have determined That the proposed algorithm compared with other online algorithms Because of sample density provided better performance Terms of accuracy. and also The number of support vectors in the our algorithm are much less than other algorithms to online classify. That in some cases the number of support vectors of 3 or more than 60 times less than other methods.

database	LIBSVM	IncrSVM	OISVM	Our method
Banana	89.75 (194)	90 (121)	89.95 (47)	90.716 (46)
Breast	78.18 (199)	78.71 (126)	75.51 (43)	77.72 (41)
Diabetes	79.21 (421)	77.83 (291)	76.83 (11)	79 (11)
Flare	67.58 (635)	68.78 (555)	67.52 (17)	69.35 (15)
German	78.49 (611)	80 (392)	77.20 (55)	78.98 (55)

database	LIBSVM	IncrSVM	OISVM	Our method
Heart	86.84 (164)	87.07 (88)	84.90 (19)	85.99 (16)
Ring norm	98.67 (477)	98 (217)	98.4 (87)	98.4 (100)
Titanic	79 (146)	77.28 (86)	77.84 (11)	78.55 (14)
Two norm	97.44 (400)	97.74 (304)	97.14 (65)	97.6 (67)
Waveform	90.46 (324)	90.3 (240)	89.67 (83)	90.2 (84)

- [6] Orabona . Francesco," On-line independent support vector machines",Pattern Recognition, Contents lists available at ScienceDirect, Pattern Recognition 43 (2010) 1402–1412
- [7] Settles Burr," Active Learning in Practice" JMLR: Workshop and Conference Proceedings 16 (2011)

#### F. Conclusion

The main challenges And proposed In the online classification data, is needed discussions about accuracy, memory and speed [7] .Since SVM for classification Data only depend on support vector Which are closer to the optimal hyperplane all other samples are irrelevant to classification action, It may reduce the accuracy of classification. In this direction we suggested an important issue to determine the density of each sample. To solve the problem of memory and speed, We have consider a set of training vectors for classification function Independent of the sample used to support vector. And used the method of Newton instead Lagrange and dual formulas. By using the proposed approach, would have considerably reduce the accuracy problem and memory requirements in online algorithm. Finally, the proposed method has been tested on 10 standard data set,also The results compared with the methods Incrsvm and OISVM and LIBSVM. we can conclude that form results of The comparison, The number of support vectors considerably reduced in this way than any other online methods. And the problem of the infinite memory requirements is solved. And also to the relative density degree of each sample, Classification provide greater accuracy, And therefore reduced the classification error.

#### REFERENCES

- [1] B.E. Boser, I.M. Guyon, V.N. Vapnik, A training algorithm for optimal margin classifiers, in: D. Haussler (Ed.), Proceedings of the 5th Annual ACMWorkshop on Computational Learning Theory, ACM Press, New York, 1992, pp. 144–152.
- [2] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods, Cambridge University Press, Cambridge, 2000.
- [3] Steinwart, Sparseness of support vector machines, Journal of Machine Learning Research 4 (2003) 1071–1105.
- [4] K. Lee, D.-W. Kim, D. Lee, K.H. Lee, Improving support vector data description using local density degree, Pattern Recognition 38 (10) (2005)
- [5] Zhang Li,"Density-induced margin support vector machines"Contents lists available at ScienceDirect journal homepage: [www.elsevier.com/locate/pr](http://www.elsevier.com/locate/pr), Pattern Recognition, Pattern Recognition 44 (2011) 1448–1460

## IJCSIS REVIEWERS' LIST

Assist Prof (Dr.) M. Emre Celebi, Louisiana State University in Shreveport, USA  
Dr. Lam Hong Lee, Universiti Tunku Abdul Rahman, Malaysia  
Dr. Shimon K. Modi, Director of Research BSPA Labs, Purdue University, USA  
Dr. Jianguo Ding, Norwegian University of Science and Technology (NTNU), Norway  
Assoc. Prof. N. Jaisankar, VIT University, Vellore, Tamilnadu, India  
Dr. Amogh Kavimandan, The Mathworks Inc., USA  
Dr. Ramasamy Mariappan, Vinayaka Missions University, India  
Dr. Yong Li, School of Electronic and Information Engineering, Beijing Jiaotong University, P.R. China  
Assist. Prof. Sugam Sharma, NIET, India / Iowa State University, USA  
Dr. Jorge A. Ruiz-Vanoye, Universidad Autónoma del Estado de Morelos, Mexico  
Dr. Neeraj Kumar, SMVD University, Katra (J&K), India  
Dr Genge Bela, "Petru Maior" University of Targu Mures, Romania  
Dr. Junjie Peng, Shanghai University, P. R. China  
Dr. Ilhem LENGILIZ, HANA Group - CRISTAL Laboratory, Tunisia  
Prof. Dr. Durgesh Kumar Mishra, Acropolis Institute of Technology and Research, Indore, MP, India  
Dr. Jorge L. Hernández-Ardieta, University Carlos III of Madrid, Spain  
Prof. Dr.C.Suresh Gnana Dhas, Anna University, India  
Dr Li Fang, Nanyang Technological University, Singapore  
Prof. Pijush Biswas, RCC Institute of Information Technology, India  
Dr. Siddhivinayak Kulkarni, University of Ballarat, Ballarat, Victoria, Australia  
Dr. A. Arul Lawrence, Royal College of Engineering & Technology, India  
Dr. Wongyos Keardsri, Chulalongkorn University, Bangkok, Thailand  
Dr. Somesh Kumar Dewangan, CSVTU Bhilai (C.G.)/ Dimat Raipur, India  
Dr. Hayder N. Jasem, University Putra Malaysia, Malaysia  
Dr. A.V.Senthil Kumar, C. M. S. College of Science and Commerce, India  
Dr. R. S. Karthik, C. M. S. College of Science and Commerce, India  
Dr. P. Vasant, University Technology Petronas, Malaysia  
Dr. Wong Kok Seng, Soongsil University, Seoul, South Korea  
Dr. Praveen Ranjan Srivastava, BITS PILANI, India  
Dr. Kong Sang Kelvin, Leong, The Hong Kong Polytechnic University, Hong Kong  
Dr. Mohd Nazri Ismail, Universiti Kuala Lumpur, Malaysia  
Dr. Rami J. Matarneh, Al-isra Private University, Amman, Jordan  
Dr Ojesanmi Olusegun Ayodeji, Ajayi Crowther University, Oyo, Nigeria  
Dr. Riktesh Srivastava, Skyline University, UAE  
Dr. Oras F. Baker, UCSI University - Kuala Lumpur, Malaysia  
Dr. Ahmed S. Ghiduk, Faculty of Science, Beni-Suef University, Egypt  
and Department of Computer science, Taif University, Saudi Arabia  
Dr. Tirthankar Gayen, IIT Kharagpur, India  
Dr. Huei-Ru Tseng, National Chiao Tung University, Taiwan

Prof. Ning Xu, Wuhan University of Technology, China  
Dr Mohammed Salem Binwahlan, Hadhramout University of Science and Technology, Yemen  
& Universiti Teknologi Malaysia, Malaysia.  
Dr. Aruna Ranganath, Bhoj Reddy Engineering College for Women, India  
Dr. Hafeezullah Amin, Institute of Information Technology, KUST, Kohat, Pakistan  
Prof. Syed S. Rizvi, University of Bridgeport, USA  
Dr. Shahbaz Pervez Chattha, University of Engineering and Technology Taxila, Pakistan  
Dr. Shishir Kumar, Jaypee University of Information Technology, Wakanaghat (HP), India  
Dr. Shahid Mumtaz, Portugal Telecommunication, Instituto de Telecomunicações (IT) , Aveiro, Portugal  
Dr. Rajesh K Shukla, Corporate Institute of Science & Technology Bhopal M P  
Dr. Poonam Garg, Institute of Management Technology, India  
Dr. S. Mehta, Inha University, Korea  
Dr. Dilip Kumar S.M, Bangalore University, Bangalore  
Prof. Malik Sikander Hayat Khiyal, Fatima Jinnah Women University, Rawalpindi, Pakistan  
Dr. Virendra Gomase , Department of Bioinformatics, Padmashree Dr. D.Y. Patil University  
Dr. Irraivan Elamvazuthi, University Technology PETRONAS, Malaysia  
Dr. Saqib Saeed, University of Siegen, Germany  
Dr. Pavan Kumar Gorakavi, IPMA-USA [YC]  
Dr. Ahmed Nabih Zaki Rashed, Menoufia University, Egypt  
Prof. Shishir K. Shandilya, Rukmani Devi Institute of Science & Technology, India  
Dr. J. Komala Lakshmi, SNR Sons College, Computer Science, India  
Dr. Muhammad Sohail, KUST, Pakistan  
Dr. Manjaiah D.H, Mangalore University, India  
Dr. S Santhosh Baboo, D.G.Vaishnav College, Chennai, India  
Prof. Dr. Mokhtar Beldjehem, Sainte-Anne University, Halifax, NS, Canada  
Dr. Deepak Laxmi Narasimha, University of Malaya, Malaysia  
Prof. Dr. Arunkumar Thangavelu, Vellore Institute Of Technology, India  
Dr. M. Azath, Anna University, India  
Dr. Md. Rabiul Islam, Rajshahi University of Engineering & Technology (RUET), Bangladesh  
Dr. Aos Alaa Zaidan Ansaef, Multimedia University, Malaysia  
Dr Suresh Jain, Devi Ahilya University, Indore (MP) India,  
Dr. Mohammed M. Kadhum, Universiti Utara Malaysia  
Dr. Hanumanthappa. J. University of Mysore, India  
Dr. Syed Ishtiaque Ahmed, Bangladesh University of Engineering and Technology (BUET)  
Dr Akinola Solomon Olalekan, University of Ibadan, Ibadan, Nigeria  
Dr. Santosh K. Pandey, The Institute of Chartered Accountants of India  
Dr. P. Vasant, Power Control Optimization, Malaysia  
Dr. Petr Ivankov, Automatika - S, Russian Federation  
Dr. Utkarsh Seetha, Data Infosys Limited, India  
Mrs. Priti Maheshwary, Maulana Azad National Institute of Technology, Bhopal  
Dr. (Mrs) Padmavathi Ganapathi, Avinashilingam University for Women, Coimbatore  
Assist. Prof. A. Neela madheswari, Anna university, India

Prof. Ganesan Ramachandra Rao, PSG College of Arts and Science, India  
Mr. Kamanashis Biswas, Daffodil International University, Bangladesh  
Dr. Atul Gonsai, Saurashtra University, Gujarat, India  
Mr. Angkoon Phinyomark, Prince of Songkla University, Thailand  
Mrs. G. Nalini Priya, Anna University, Chennai  
Dr. P. Subashini, Avinashilingam University for Women, India  
Assoc. Prof. Vijay Kumar Chakka, Dhirubhai Ambani IICT, Gandhinagar, Gujarat  
Mr Jitendra Agrawal, : Rajiv Gandhi Proudhyogiki Vishwavidyalaya, Bhopal  
Mr. Vishal Goyal, Department of Computer Science, Punjabi University, India  
Dr. R. Baskaran, Department of Computer Science and Engineering, Anna University, Chennai  
Assist. Prof, Kanwalvir Singh Dhindsa, B.B.S.B.Engg.College, Fatehgarh Sahib (Punjab), India  
Dr. Jamal Ahmad Dargham, School of Engineering and Information Technology, Universiti Malaysia Sabah  
Mr. Nitin Bhatia, DAV College, India  
Dr. Dhavachelvan Ponnuram, Pondicherry Central University, India  
Dr. Mohd Faizal Abdollah, University of Technical Malaysia, Malaysia  
Assist. Prof. Sonal Chawla, Panjab University, India  
Dr. Abdul Wahid, AKG Engg. College, Ghaziabad, India  
Mr. Arash Habibi Lashkari, University of Malaya (UM), Malaysia  
Mr. Md. Rajibul Islam, Ibnu Sina Institute, University Technology Malaysia  
Professor Dr. Sabu M. Thampi, .B.S Institute of Technology for Women, Kerala University, India  
Mr. Noor Muhammed Nayeem, Université Lumière Lyon 2, 69007 Lyon, France  
Dr. Himanshu Aggarwal, Department of Computer Engineering, Punjabi University, India  
Prof R. Naidoo, Dept of Mathematics/Center for Advanced Computer Modelling, Durban University of  
Technology, Durban, South Africa  
Prof. Mydhili K Nair, Visweswaraiyah Technological University, Bangalore, India  
M. Prabu, Adhiyamaan College of Engineering/Anna University, India  
Mr. Swakkhar Shatabda, United International University, Bangladesh  
Dr. Abdur Rashid Khan, ICIT, Gomal University, Dera Ismail Khan, Pakistan  
Mr. H. Abdul Shabeer, I-Nautix Technologies, Chennai, India  
Dr. M. Aramudhan, Perunthalaivar Kamarajar Institute of Engineering and Technology, India  
Dr. M. P. Thapliyal, Department of Computer Science, HNB Garhwal University (Central University), India  
Dr. Shahaboddin Shamshirband, Islamic Azad University, Iran  
Mr. Zeashan Hameed Khan, Université de Grenoble, France  
Prof. Anil K Ahlawat, Ajay Kumar Garg Engineering College, Ghaziabad, UP Technical University, Lucknow  
Mr. Longe Olumide Babatope, University Of Ibadan, Nigeria  
Associate Prof. Raman Maini, University College of Engineering, Punjabi University, India  
Dr. Maslin Masrom, University Technology Malaysia, Malaysia  
Sudipta Chattopadhyay, Jadavpur University, Kolkata, India  
Dr. Dang Tuan NGUYEN, University of Information Technology, Vietnam National University - Ho Chi Minh  
City  
Dr. Mary Lourde R., BITS-PILANI Dubai, UAE  
Dr. Abdul Aziz, University of Central Punjab, Pakistan

Mr. Karan Singh, Gautam Budtha University, India  
Mr. Avinash Pokhriyal, Uttar Pradesh Technical University, Lucknow, India  
Associate Prof Dr Zuraini Ismail, University Technology Malaysia, Malaysia  
Assistant Prof. Yasser M. Alginahi, Taibah University, Madinah Munawwarah, KSA  
Mr. Dakshina Ranjan Kisku, West Bengal University of Technology, India  
Mr. Raman Kumar, Dr B R Ambedkar National Institute of Technology, Jalandhar, Punjab, India  
Associate Prof. Samir B. Patel, Institute of Technology, Nirma University, India  
Dr. M.Munir Ahamed Rabbani, B. S. Abdur Rahman University, India  
Asst. Prof. Koushik Majumder, West Bengal University of Technology, India  
Dr. Alex Pappachen James, Queensland Micro-nanotechnology center, Griffith University, Australia  
Assistant Prof. S. Hariharan, B.S. Abdur Rahman University, India  
Asst Prof. Jasmine. K. S, R.V.College of Engineering, India  
Mr Naushad Ali Mamode Khan, Ministry of Education and Human Resources, Mauritius  
Prof. Mahesh Goyani, G H Patel Collge of Engg. & Tech, V.V.N, Anand, Gujarat, India  
Dr. Mana Mohammed, University of Tlemcen, Algeria  
Prof. Jatinder Singh, Universal Institutiion of Engg. & Tech. CHD, India  
Mrs. M. Anandhavalli Gauthaman, Sikkim Manipal Institute of Technology, Majitar, East Sikkim  
Dr. Bin Guo, Institute Telecom SudParis, France  
Mrs. Maleika Mehr Nigar Mohamed Heenaye-Mamode Khan, University of Mauritius  
Prof. Pijush Biswas, RCC Institute of Information Technology, India  
Mr. V. Bala Dhandayuthapani, Mekelle University, Ethiopia  
Dr. Irfan Syamsuddin, State Polytechnic of Ujung Pandang, Indonesia  
Mr. Kavi Kumar Khedo, University of Mauritius, Mauritius  
Mr. Ravi Chandiran, Zagro Singapore Pte Ltd. Singapore  
Mr. Milindkumar V. Sarode, Jawaharlal Darda Institute of Engineering and Technology, India  
Dr. Shamimul Qamar, KSJ Institute of Engineering & Technology, India  
Dr. C. Arun, Anna University, India  
Assist. Prof. M.N.Birje, Basaveshwar Engineering College, India  
Prof. Hamid Reza Naji, Department of Computer Enigneering, Shahid Beheshti University, Tehran, Iran  
Assist. Prof. Debasis Giri, Department of Computer Science and Engineering, Haldia Institute of Technology  
Subhabrata Barman, Haldia Institute of Technology, West Bengal  
Mr. M. I. Lali, COMSATS Institute of Information Technology, Islamabad, Pakistan  
Dr. Feroz Khan, Central Institute of Medicinal and Aromatic Plants, Lucknow, India  
Mr. R. Nagendran, Institute of Technology, Coimbatore, Tamilnadu, India  
Mr. Amnach Khawne, King Mongkut's Institute of Technology Ladkrabang, Ladkrabang, Bangkok, Thailand  
Dr. P. Chakrabarti, Sir Padampat Singhanian University, Udaipur, India  
Mr. Nafiz Imtiaz Bin Hamid, Islamic University of Technology (IUT), Bangladesh.  
Shahab-A. Shamshirband, Islamic Azad University, Chalous, Iran  
Prof. B. Priestly Shan, Anna Univeristy, Tamilnadu, India  
Venkatramreddy Velma, Dept. of Bioinformatics, University of Mississippi Medical Center, Jackson MS USA  
Akshi Kumar, Dept. of Computer Engineering, Delhi Technological University, India  
Dr. Umesh Kumar Singh, Vikram University, Ujjain, India



Mr. Serguei A. Mokhov, Concordia University, Canada  
Mr. Lai Khin Wee, Universiti Teknologi Malaysia, Malaysia  
Dr. Awadhesh Kumar Sharma, Madan Mohan Malviya Engineering College, India  
Mr. Syed R. Rizvi, Analytical Services & Materials, Inc., USA  
Dr. S. Karthik, SNS College of Technology, India  
Mr. Syed Qasim Bukhari, CIMET (Universidad de Granada), Spain  
Mr. A.D.Potgantwar, Pune University, India  
Dr. Himanshu Aggarwal, Punjabi University, India  
Mr. Rajesh Ramachandran, Naipunya Institute of Management and Information Technology, India  
Dr. K.L. Shunmuganathan, R.M.K Engg College , Kavaraipettai ,Chennai  
Dr. Prasant Kumar Pattnaik, KIST, India.  
Dr. Ch. Aswani Kumar, VIT University, India  
Mr. Ijaz Ali Shoukat, King Saud University, Riyadh KSA  
Mr. Arun Kumar, Sir Padam Pat Singhanian University, Udaipur, Rajasthan  
Mr. Muhammad Imran Khan, Universiti Teknologi PETRONAS, Malaysia  
Dr. Natarajan Meghanathan, Jackson State University, Jackson, MS, USA  
Mr. Mohd Zaki Bin Mas'ud, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia  
Prof. Dr. R. Geetharamani, Dept. of Computer Science and Eng., Rajalakshmi Engineering College, India  
Dr. Smita Rajpal, Institute of Technology and Management, Gurgaon, India  
Dr. S. Abdul Khader Jilani, University of Tabuk, Tabuk, Saudi Arabia  
Mr. Syed Jamal Haider Zaidi, Bahria University, Pakistan  
Dr. N. Devarajan, Government College of Technology, Coimbatore, Tamilnadu, INDIA  
Mr. R. Jagadeesh Kannan, RMK Engineering College, India  
Mr. Deo Prakash, Shri Mata Vaishno Devi University, India  
Mr. Mohammad Abu Naser, Dept. of EEE, IUT, Gazipur, Bangladesh  
Assist. Prof. Prasun Ghosal, Bengal Engineering and Science University, India  
Mr. Md. Golam Kaosar, School of Engineering and Science, Victoria University, Melbourne City, Australia  
Mr. R. Mahammad Shafi, Madanapalle Institute of Technology & Science, India  
Dr. F.Sagayaraj Francis, Pondicherry Engineering College, India  
Dr. Ajay Goel, HIET , Kaithal, India  
Mr. Nayak Sunil Kashibarao, Bahirji Smarak Mahavidyalaya, India  
Mr. Suhas J Manangi, Microsoft India  
Dr. Kalyankar N. V., Yeshwant Mahavidyalaya, Nanded , India  
Dr. K.D. Verma, S.V. College of Post graduate studies & Research, India  
Dr. Amjad Rehman, University Technology Malaysia, Malaysia  
Mr. Rachit Garg, L K College, Jalandhar, Punjab  
Mr. J. William, M.A.M college of Engineering, Trichy, Tamilnadu, India  
Prof. Jue-Sam Chou, Nanhua University, College of Science and Technology, Taiwan  
Dr. Thorat S.B., Institute of Technology and Management, India  
Mr. Ajay Prasad, Sir Padampat Singhanian University, Udaipur, India  
Dr. Kamaljit I. Lakhtaria, Atmiya Institute of Technology & Science, India  
Mr. Syed Rafiul Hussain, Ahsanullah University of Science and Technology, Bangladesh

Mrs Fazeela Tunnisa, Najran University, Kingdom of Saudi Arabia  
Mrs Kavita Taneja, Maharishi Markandeshwar University, Haryana, India  
Mr. Maniyar Shiraz Ahmed, Najran University, Najran, KSA  
Mr. Anand Kumar, AMC Engineering College, Bangalore  
Dr. Rakesh Chandra Gangwar, Beant College of Engg. & Tech., Gurdaspur (Punjab) India  
Dr. V V Rama Prasad, Sree Vidyanikethan Engineering College, India  
Assist. Prof. Neetesh Kumar Gupta, Technocrats Institute of Technology, Bhopal (M.P.), India  
Mr. Ashish Seth, Uttar Pradesh Technical University, Lucknow ,UP India  
Dr. V V S S S Balaram, Sreenidhi Institute of Science and Technology, India  
Mr Rahul Bhatia, Lingaya's Institute of Management and Technology, India  
Prof. Niranjana Reddy. P, KITS , Warangal, India  
Prof. Rakesh. Lingappa, Vijetha Institute of Technology, Bangalore, India  
Dr. Mohammed Ali Hussain, Nimra College of Engineering & Technology, Vijayawada, A.P., India  
Dr. A.Srinivasan, MNM Jain Engineering College, Rajiv Gandhi Salai, Thorapakkam, Chennai  
Mr. Rakesh Kumar, M.M. University, Mullana, Ambala, India  
Dr. Lena Khaled, Zarqa Private University, Aman, Jordan  
Ms. Supriya Kapoor, Patni/Lingaya's Institute of Management and Tech., India  
Dr. Tossapon Boongoen , Aberystwyth University, UK  
Dr . Bilal Alatas, Firat University, Turkey  
Assist. Prof. Jyoti Praaksh Singh , Academy of Technology, India  
Dr. Ritu Soni, GNG College, India  
Dr . Mahendra Kumar , Sagar Institute of Research & Technology, Bhopal, India.  
Dr. Binod Kumar, Lakshmi Narayan College of Tech.(LNCT)Bhopal India  
Dr. Muzhir Shaban Al-Ani, Amman Arab University Amman – Jordan  
Dr. T.C. Manjunath , ATRIA Institute of Tech, India  
Mr. Muhammad Zakarya, COMSATS Institute of Information Technology (CIIT), Pakistan  
Assist. Prof. Harmunish Taneja, M. M. University, India  
Dr. Chitra Dhawale , SICSR, Model Colony, Pune, India  
Mrs Sankari Muthukaruppan, Nehru Institute of Engineering and Technology, Anna University, India  
Mr. Aaqif Afzaal Abbasi, National University Of Sciences And Technology, Islamabad  
Prof. Ashutosh Kumar Dubey, Trinity Institute of Technology and Research Bhopal, India  
Mr. G. Appasami, Dr. Pauls Engineering College, India  
Mr. M Yasin, National University of Science and Tech, karachi (NUST), Pakistan  
Mr. Yaser Miaji, University Utara Malaysia, Malaysia  
Mr. Shah Ahsanul Haque, International Islamic University Chittagong (IIUC), Bangladesh  
Prof. (Dr) Syed Abdul Sattar, Royal Institute of Technology & Science, India  
Dr. S. Sasikumar, Roever Engineering College  
Assist. Prof. Monit Kapoor, Maharishi Markandeshwar University, India  
Mr. Nwaocha Vivian O, National Open University of Nigeria  
Dr. M. S. Vijaya, GR Govindarajulu School of Applied Computer Technology, India  
Assist. Prof. Chakresh Kumar, Manav Rachna International University, India  
Mr. Kunal Chadha , R&D Software Engineer, Gemalto, Singapore

Mr. Mueen Uddin, Universiti Teknologi Malaysia, UTM , Malaysia  
Dr. Dhuha Basheer abdullah, Mosul university, Iraq  
Mr. S. Audithan, Annamalai University, India  
Prof. Vijay K Chaudhari, Technocrats Institute of Technology , India  
Associate Prof. Mohd Ilyas Khan, Technocrats Institute of Technology , India  
Dr. Vu Thanh Nguyen, University of Information Technology, HoChiMinh City, VietNam  
Assist. Prof. Anand Sharma, MITS, Lakshmangarh, Sikar, Rajasthan, India  
Prof. T V Narayana Rao, HITAM Engineering college, Hyderabad  
Mr. Deepak Gour, Sir Padampat Singhania University, India  
Assist. Prof. Amutharaj Joyson, Kalasalingam University, India  
Mr. Ali Balador, Islamic Azad University, Iran  
Mr. Mohit Jain, Maharaja Surajmal Institute of Technology, India  
Mr. Dilip Kumar Sharma, GLA Institute of Technology & Management, India  
Dr. Debojyoti Mitra, Sir padampat Singhania University, India  
Dr. Ali Dehghantanha, Asia-Pacific University College of Technology and Innovation, Malaysia  
Mr. Zhao Zhang, City University of Hong Kong, China  
Prof. S.P. Setty, A.U. College of Engineering, India  
Prof. Patel Rakeshkumar Kantilal, Sankalchand Patel College of Engineering, India  
Mr. Biswajit Bhowmik, Bengal College of Engineering & Technology, India  
Mr. Manoj Gupta, Apex Institute of Engineering & Technology, India  
Assist. Prof. Ajay Sharma, Raj Kumar Goel Institute Of Technology, India  
Assist. Prof. Ramveer Singh, Raj Kumar Goel Institute of Technology, India  
Dr. Hanan Elazhary, Electronics Research Institute, Egypt  
Dr. Hosam I. Faiq, USM, Malaysia  
Prof. Dipti D. Patil, MAEER's MIT College of Engg. & Tech, Pune, India  
Assist. Prof. Devendra Chack, BCT Kumaon engineering College Dwarahat Almora, India  
Prof. Manpreet Singh, M. M. Engg. College, M. M. University, India  
Assist. Prof. M. Sadiq ali Khan, University of Karachi, Pakistan  
Mr. Prasad S. Halgaonkar, MIT - College of Engineering, Pune, India  
Dr. Imran Ghani, Universiti Teknologi Malaysia, Malaysia  
Prof. Varun Kumar Kakar, Kumaon Engineering College, Dwarahat, India  
Assist. Prof. Nisheeth Joshi, Apaji Institute, Banasthali University, Rajasthan, India  
Associate Prof. Kunwar S. Vaisla, VCT Kumaon Engineering College, India  
Prof Anupam Choudhary, Bhilai School Of Engg.,Bhilai (C.G.),India  
Mr. Divya Prakash Shrivastava, Al Jabal Al garbi University, Zawya, Libya  
Associate Prof. Dr. V. Radha, Avinashilingam Deemed university for women, Coimbatore.  
Dr. Kasarapu Ramani, JNT University, Anantapur, India  
Dr. Anuraag Awasthi, Jayoti Vidyapeeth Womens University, India  
Dr. C G Ravichandran, R V S College of Engineering and Technology, India  
Dr. Mohamed A. Deriche, King Fahd University of Petroleum and Minerals, Saudi Arabia  
Mr. Abbas Karimi, Universiti Putra Malaysia, Malaysia  
Mr. Amit Kumar, Jaypee University of Engg. and Tech., India

Dr. Nikolai Stoianov, Defense Institute, Bulgaria  
Assist. Prof. S. Ranichandra, KSR College of Arts and Science, Tiruchencode  
Mr. T.K.P. Rajagopal, Diamond Horse International Pvt Ltd, India  
Dr. Md. Ekramul Hamid, Rajshahi University, Bangladesh  
Mr. Hemanta Kumar Kalita , TATA Consultancy Services (TCS), India  
Dr. Messaouda Azzouzi, Ziane Achour University of Djelfa, Algeria  
Prof. (Dr.) Juan Jose Martinez Castillo, "Gran Mariscal de Ayacucho" University and Acantelys research Group, Venezuela  
Dr. Jatinderkumar R. Saini, Narmada College of Computer Application, India  
Dr. Babak Bashari Rad, University Technology of Malaysia, Malaysia  
Dr. Nighat Mir, Effat University, Saudi Arabia  
Prof. (Dr.) G.M.Nasira, Sasurie College of Engineering, India  
Mr. Varun Mittal, Gemalto Pte Ltd, Singapore  
Assist. Prof. Mrs P. Banumathi, Kathir College Of Engineering, Coimbatore  
Assist. Prof. Quan Yuan, University of Wisconsin-Stevens Point, US  
Dr. Pranam Paul, Narula Institute of Technology, Agarpara, West Bengal, India  
Assist. Prof. J. Ramkumar, V.L.B Janakiammal college of Arts & Science, India  
Mr. P. Sivakumar, Anna university, Chennai, India  
Mr. Md. Humayun Kabir Biswas, King Khalid University, Kingdom of Saudi Arabia  
Mr. Mayank Singh, J.P. Institute of Engg & Technology, Meerut, India  
HJ. Kamaruzaman Jusoff, Universiti Putra Malaysia  
Mr. Nikhil Patrick Lobo, CADES, India  
Dr. Amit Wason, Rayat-Bahra Institute of Engineering & Boi-Technology, India  
Dr. Rajesh Shrivastava, Govt. Benazir Science & Commerce College, Bhopal, India  
Assist. Prof. Vishal Bharti, DCE, Gurgaon  
Mrs. Sunita Bansal, Birla Institute of Technology & Science, India  
Dr. R. Sudhakar, Dr.Mahalingam college of Engineering and Technology, India  
Dr. Amit Kumar Garg, Shri Mata Vaishno Devi University, Katra(J&K), India  
Assist. Prof. Raj Gaurang Tiwari, AZAD Institute of Engineering and Technology, India  
Mr. Hamed Taherdoost, Tehran, Iran  
Mr. Amin Daneshmand Malayeri, YRC, IAU, Malayer Branch, Iran  
Mr. Shantanu Pal, University of Calcutta, India  
Dr. Terry H. Walcott, E-Promag Consultancy Group, United Kingdom  
Dr. Ezekiel U OKIKE, University of Ibadan, Nigeria  
Mr. P. Mahalingam, Caledonian College of Engineering, Oman  
Dr. Mahmoud M. A. Abd Ellatif, Mansoura University, Egypt  
Prof. Kunwar S. Vaisla, BCT Kumaon Engineering College, India  
Prof. Mahesh H. Panchal, Kalol Institute of Technology & Research Centre, India  
Mr. Muhammad Asad, Technical University of Munich, Germany  
Mr. AliReza Shams Shafigh, Azad Islamic university, Iran  
Prof. S. V. Nagaraj, RMK Engineering College, India  
Mr. Ashikali M Hasan, Senior Researcher, CeINet security, India

Dr. Adnan Shahid Khan, University Technology Malaysia, Malaysia  
Mr. Prakash Gajanan Burade, Nagpur University/ITM college of engg, Nagpur, India  
Dr. Jagdish B.Helonde, Nagpur University/ITM college of engg, Nagpur, India  
Professor, Doctor BOUHORMA Mohammed, Univertsity Abdelmalek Essaadi, Morocco  
Mr. K. Thirumalaivasan, Pondicherry Engg. College, India  
Mr. Umbarkar Anantkumar Janardan, Walchand College of Engineering, India  
Mr. Ashish Chaurasia, Gyan Ganga Institute of Technology & Sciences, India  
Mr. Sunil Taneja, Kurukshetra University, India  
Mr. Fauzi Adi Rafrastara, Dian Nuswantoro University, Indonesia  
Dr. Yaduvir Singh, Thapar University, India  
Dr. Ioannis V. Koskosas, University of Western Macedonia, Greece  
Dr. Vasantha Kalyani David, Avinashilingam University for women, Coimbatore  
Dr. Ahmed Mansour Manasrah, Universiti Sains Malaysia, Malaysia  
Miss. Nazanin Sadat Kazazi, University Technology Malaysia, Malaysia  
Mr. Saeed Rasouli Heikalabad, Islamic Azad University - Tabriz Branch, Iran  
Assoc. Prof. Dharendra Mishra, SVKM's NMIMS University, India  
Prof. Shapoor Zarei, UAE Inventors Association, UAE  
Prof. B.Raja Sarath Kumar, Lenora College of Engineering, India  
Dr. Bashir Alam, Jamia millia Islamia, Delhi, India  
Prof. Anant J Umbarkar, Walchand College of Engg., India  
Assist. Prof. B. Bharathi, Sathyabama University, India  
Dr. Fokrul Alom Mazarbhuiya, King Khalid University, Saudi Arabia  
Prof. T.S.Jeyali Laseeth, Anna University of Technology, Tirunelveli, India  
Dr. M. Balraju, Jawahar Lal Nehru Technological University Hyderabad, India  
Dr. Vijayalakshmi M. N., R.V.College of Engineering, Bangalore  
Prof. Walid Moudani, Lebanese University, Lebanon  
Dr. Saurabh Pal, VBS Purvanchal University, Jaunpur, India  
Associate Prof. Suneet Chaudhary, Dehradun Institute of Technology, India  
Associate Prof. Dr. Manuj Darbari, BBD University, India  
Ms. Prema Selvaraj, K.S.R College of Arts and Science, India  
Assist. Prof. Ms.S.Sasikala, KSR College of Arts & Science, India  
Mr. Sukhvinder Singh Deora, NC Institute of Computer Sciences, India  
Dr. Abhay Bansal, Amity School of Engineering & Technology, India  
Ms. Sumita Mishra, Amity School of Engineering and Technology, India  
Professor S. Viswanadha Raju, JNT University Hyderabad, India  
Mr. Asghar Shahrzad Khashandarag, Islamic Azad University Tabriz Branch, India  
Mr. Manoj Sharma, Panipat Institute of Engg. & Technology, India  
Mr. Shakeel Ahmed, King Faisal University, Saudi Arabia  
Dr. Mohamed Ali Mahjoub, Institute of Engineer of Monastir, Tunisia  
Mr. Adri Jovin J.J., SriGuru Institute of Technology, India  
Dr. Sukumar Senthilkumar, Universiti Sains Malaysia, Malaysia  
Mr. Rakesh Bharati, Dehradun Institute of Technology Dehradun, India

Mr. Shervan Fekri Ershad, Shiraz International University, Iran  
Mr. Md. Safiqul Islam, Daffodil International University, Bangladesh  
Mr. Mahmudul Hasan, Daffodil International University, Bangladesh  
Prof. Mandakini Tayade, UIT, RGTU, Bhopal, India  
Ms. Sarla More, UIT, RGTU, Bhopal, India  
Mr. Tushar Hrishikesh Jaware, R.C. Patel Institute of Technology, Shirpur, India  
Ms. C. Divya, Dr G R Damodaran College of Science, Coimbatore, India  
Mr. Fahimuddin Shaik, Annamacharya Institute of Technology & Sciences, India  
Dr. M. N. Giri Prasad, JNTUCE, Pulivendula, A.P., India  
Assist. Prof. Chintan M Bhatt, Charotar University of Science And Technology, India  
Prof. Sahista Machchhar, Marwadi Education Foundation's Group of institutions, India  
Assist. Prof. Navnish Goel, S. D. College Of Enginnering & Technology, India  
Mr. Khaja Kamaluddin, Sirt University, Sirt, Libya  
Mr. Mohammad Zaidul Karim, Daffodil International, Bangladesh  
Mr. M. Vijayakumar, KSR College of Engineering, Tiruchengode, India  
Mr. S. A. Ahsan Rajon, Khulna University, Bangladesh  
Dr. Muhammad Mohsin Nazir, LCW University Lahore, Pakistan  
Mr. Mohammad Asadul Hoque, University of Alabama, USA  
Mr. P.V.Sarathchand, Indur Institute of Engineering and Technology, India  
Mr. Durgesh Samadhiya, Chung Hua University, Taiwan  
Dr Venu Kuthadi, University of Johannesburg, Johannesburg, RSA  
Dr. (Er) Jasvir Singh, Guru Nanak Dev University, Amritsar, Punjab, India  
Mr. Jasmin Cosic, Min. of the Interior of Una-sana canton, B&H, Bosnia and Herzegovina  
Dr S. Rajalakshmi, Botho College, South Africa  
Dr. Mohamed Sarrab, De Montfort University, UK  
Mr. Basappa B. Kodada, Canara Engineering College, India  
Assist. Prof. K. Ramana, Annamacharya Institute of Technology and Sciences, India  
Dr. Ashu Gupta, Apeejay Institute of Management, Jalandhar, India  
Assist. Prof. Shaik Rasool, Shadan College of Engineering & Technology, India  
Assist. Prof. K. Suresh, Annamacharya Institute of Tech & Sci. Rajampet, AP, India  
Dr . G. Singaravel, K.S.R. College of Engineering, India  
Dr B. G. Geetha, K.S.R. College of Engineering, India  
Assist. Prof. Kavita Choudhary, ITM University, Gurgaon  
Dr. Mehrdad Jalali, Azad University, Mashhad, Iran  
Megha Goel, Shamli Institute of Engineering and Technology, Shamli, India  
Mr. Chi-Hua Chen, Institute of Information Management, National Chiao-Tung University, Taiwan (R.O.C.)  
Assoc. Prof. A. Rajendran, RVS College of Engineering and Technology, India  
Assist. Prof. S. Jaganathan, RVS College of Engineering and Technology, India  
Assoc. Prof. (Dr.) A S N Chakravarthy, JNTUK University College of Engineering Vizianagaram (State University)  
Assist. Prof. Deepshikha Patel, Technocrat Institute of Technology, India  
Assist. Prof. Maram Balajee, GMRIT, India

Assist. Prof. Monika Bhatnagar, TIT, India  
Prof. Gaurang Panchal, Charotar University of Science & Technology, India  
Prof. Anand K. Tripathi, Computer Society of India  
Prof. Jyoti Chaudhary, High Performance Computing Research Lab, India  
Assist. Prof. Supriya Raheja, ITM University, India  
Dr. Pankaj Gupta, Microsoft Corporation, U.S.A.  
Assist. Prof. Panchamukesh Chandaka, Hyderabad Institute of Tech. & Management, India  
Prof. Mohan H.S, SJB Institute Of Technology, India  
Mr. Hossein Malekinezhad, Islamic Azad University, Iran  
Mr. Zatin Gupta, Universti Malaysia, Malaysia  
Assist. Prof. Amit Chauhan, Phonics Group of Institutions, India  
Assist. Prof. Ajal A. J., METS School Of Engineering, India  
Mrs. Omowunmi Omobola Adeyemo, University of Ibadan, Nigeria  
Dr. Bharat Bhushan Agarwal, I.F.T.M. University, India  
Md. Nazrul Islam, University of Western Ontario, Canada  
Tushar Kanti, L.N.C.T, Bhopal, India  
Er. Aumreesh Kumar Saxena, SIRTs College Bhopal, India  
Mr. Mohammad Monirul Islam, Daffodil International University, Bangladesh  
Dr. Kashif Nisar, University Utara Malaysia, Malaysia  
Dr. Wei Zheng, Rutgers Univ/ A10 Networks, USA  
Associate Prof. Rituraj Jain, Vyas Institute of Engg & Tech, Jodhpur – Rajasthan  
Assist. Prof. Apoorvi Sood, I.T.M. University, India  
Dr. Kayhan Zrar Ghafoor, University Technology Malaysia, Malaysia  
Mr. Swapnil Soner, Truba Institute College of Engineering & Technology, Indore, India  
Ms. Yogita Gigras, I.T.M. University, India  
Associate Prof. Neelima Sadineni, Pydha Engineering College, India Pydha Engineering College  
Assist. Prof. K. Deepika Rani, HITAM, Hyderabad  
Ms. Shikha Maheshwari, Jaipur Engineering College & Research Centre, India  
Prof. Dr V S Giridhar Akula, Avanthi's Scientific Tech. & Research Academy, Hyderabad  
Prof. Dr.S.Saravanan, Muthayammal Engineering College, India  
Mr. Mehdi Golsorkhatabar Amiri, Islamic Azad University, Iran  
Prof. Amit Sadanand Savyanavar, MITCOE, Pune, India  
Assist. Prof. P.Oliver Jayaprakash, Anna University, Chennai  
Assist. Prof. Ms. Sujata, ITM University, Gurgaon, India  
Dr. Asoke Nath, St. Xavier's College, India  
Mr. Masoud Rafighi, Islamic Azad University, Iran  
Assist. Prof. RamBabu Pemula, NIMRA College of Engineering & Technology, India  
Assist. Prof. Ms Rita Chhikara, ITM University, Gurgaon, India  
Mr. Sandeep Maan, Government Post Graduate College, India  
Prof. Dr. S. Muralidharan, Mepco Schlenk Engineering College, India  
Associate Prof. T.V.Sai Krishna, QIS College of Engineering and Technology, India  
Mr. R. Balu, Bharathiar University, Coimbatore, India



Assist. Prof. Shekhar. R, Dr.SM College of Engineering, India  
Prof. P. Senthilkumar, Vivekanandha Institute of Engineering and Technology for Woman, India  
Mr. M. Kamarajan, PSNA College of Engineering & Technology, India  
Dr. Angajala Srinivasa Rao, Jawaharlal Nehru Technical University, India  
Assist. Prof. C. Venkatesh, A.I.T.S, Rajampet, India  
Mr. Afshin Rezakhani Roozbahani, Ayatollah Boroujerdi University, Iran  
Mr. Laxmi chand, SCTL, Noida, India  
Dr. Dr. Abdul Hannan, Vivekanand College, Aurangabad  
Prof. Mahesh Panchal, KITRC, Gujarat  
Dr. A. Subramani, K.S.R. College of Engineering, Tiruchengode  
Assist. Prof. Prakash M, Rajalakshmi Engineering College, Chennai, India  
Assist. Prof. Akhilesh K Sharma, Sir Padampat Singhania University, India  
Ms. Varsha Sahni, Guru Nanak Dev Engineering College, Ludhiana, India  
Associate Prof. Trilochan Rout, NM Institute of Engineering and Technology, India  
Mr. Srikanta Kumar Mohapatra, NMIET, Orissa, India  
Mr. Waqas Haider Bangyal, Iqra University Islamabad, Pakistan  
Dr. S. Vijayaragavan, Christ College of Engineering and Technology, Pondicherry, India  
Prof. Elboukhari Mohamed, University Mohammed First, Oujda, Morocco  
Dr. Muhammad Asif Khan, King Faisal University, Saudi Arabia  
Dr. Nagy Ramadan Darwish Omran, Cairo University, Egypt.  
Assistant Prof. Anand Nayyar, KCL Institute of Management and Technology, India  
Mr. G. Premsankar, Ericsson, India  
Assist. Prof. T. Hemalatha, VELS University, India  
Prof. Tejaswini Apte, University of Pune, India  
Dr. Edmund Ng Giap Weng, Universiti Malaysia Sarawak, Malaysia  
Mr. Mahdi Nouri, Iran University of Science and Technology, Iran  
Associate Prof. S. Asif Hussain, Annamacharya Institute of technology & Sciences, India  
Mrs. Kavita Pabreja, Maharaja Surajmal Institute (an affiliate of GGSIP University), India  
Mr. Vorugunti Chandra Sekhar, DA-IICT, India  
Mr. Muhammad Najmi Ahmad Zabidi, Universiti Teknologi Malaysia, Malaysia  
Dr. Aderemi A. Atayero, Covenant University, Nigeria  
Assist. Prof. Osama Sohaib, Balochistan University of Information Technology, Pakistan  
Assist. Prof. K. Suresh, Annamacharya Institute of Technology and Sciences, India  
Mr. Hassen Mohammed Abdulllah Alsafi, International Islamic University Malaysia (IIUM) Malaysia  
Mr. Robail Yasrab, Virtual University of Pakistan, Pakistan  
Mr. R. Balu, Bharathiar University, Coimbatore, India  
Prof. Anand Nayyar, KCL Institute of Management and Technology, Jalandhar  
Assoc. Prof. Vivek S Deshpande, MIT College of Engineering, India  
Prof. K. Saravanan, Anna university Coimbatore, India  
Dr. Ravendra Singh, MJP Rohilkhand University, Bareilly, India  
Mr. V. Mathivanan, IBRA College of Technology, Sultanate of OMAN  
Assoc. Prof. S. Asif Hussain, AITS, India

Assist. Prof. C. Venkatesh, AITS, India  
Mr. Sami Ulhaq, SZABIST Islamabad, Pakistan  
Dr. B. Justus Rabi, Institute of Science & Technology, India  
Mr. Anuj Kumar Yadav, Dehradun Institute of technology, India  
Mr. Alejandro Mosquera, University of Alicante, Spain  
Assist. Prof. Arjun Singh, Sir Padampat Singhania University (SPSU), Udaipur, India  
Dr. Smriti Agrawal, JB Institute of Engineering and Technology, Hyderabad  
Assist. Prof. Swathi Sambangi, Visakha Institute of Engineering and Technology, India  
Ms. Prabhjot Kaur, Guru Gobind Singh Indraprastha University, India  
Mrs. Samaher AL-Hothali, Yanbu University College, Saudi Arabia  
Prof. Rajneeshkaur Bedi, MIT College of Engineering, Pune, India  
Mr. Hassen Mohammed Abdullallah Alsafi, International Islamic University Malaysia (IIUM)  
Dr. Wei Zhang, Amazon.com, Seattle, WA, USA  
Mr. B. Santhosh Kumar, C S I College of Engineering, Tamil Nadu  
Dr. K. Reji Kumar, , N S S College, Pandalam, India  
Assoc. Prof. K. Seshadri Sastry, EILM University, India  
Mr. Kai Pan, UNC Charlotte, USA  
Mr. Ruikar Sachin, SGGSIET, India  
Prof. (Dr.) Vinodani Katiyar, Sri Ramswaroop Memorial University, India  
Assoc. Prof., M. Giri, Sreenivasa Institute of Technology and Management Studies, India  
Assoc. Prof. Labib Francis Gergis, Misr Academy for Engineering and Technology (MET), Egypt  
Assist. Prof. Amanpreet Kaur, ITM University, India  
Assist. Prof. Anand Singh Rajawat, Shri Vaishnav Institute of Technology & Science, Indore  
Mrs. Hadeel Saleh Haj Aliwi, Universiti Sains Malaysia (USM), Malaysia  
Dr. Abhay Bansal, Amity University, India  
Dr. Mohammad A. Mezher, Fahad Bin Sultan University, KSA  
Assist. Prof. Nidhi Arora, M.C.A. Institute, India  
Prof. Dr. P. Suresh, Karpagam College of Engineering, Coimbatore, India  
Dr. Kannan Balasubramanian, Mepco Schlenk Engineering College, India  
Dr. S. Sankara Gomathi, Panimalar Engineering college, India  
Prof. Anil kumar Suthar, Gujarat Technological University, L.C. Institute of Technology, India  
Assist. Prof. R. Hubert Rajan, NOORUL ISLAM UNIVERSITY, India  
Assist. Prof. Dr. Jyoti Mahajan, College of Engineering & Technology  
Assist. Prof. Homam Reda El-Taj, College of Network Engineering, Saudi Arabia & Malaysia  
Mr. Bijan Paul, Shahjalal University of Science & Technology, Bangladesh  
Assoc. Prof. Dr. Ch V Phani Krishna, KL University, India  
Dr. Vishal Bhatnagar, Ambedkar Institute of Advanced Communication Technologies & Research, India  
Dr. Lamri LAOUAMER, Al Qassim University, Dept. Info. Systems & European University of Brittany, Dept. Computer Science, UBO, Brest, France  
Prof. Ashish Babanrao Sasankar, G.H.Raisoni Institute Of Information Technology, India  
Prof. Pawan Kumar Goel, Shamli Institute of Engineering and Technology, India  
Mr. Ram Kumar Singh, S.V Subharti University, India

Assistant Prof. Sunish Kumar O S, Amalijothe College of Engineering, India  
Dr Sanjay Bhargava, Banasthali University, India  
Mr. Pankaj S. Kulkarni, AVEW's Shatabdi Institute of Technology, India  
Mr. Roohollah Etemadi, Islamic Azad University, Iran  
Mr. Oloruntoyin Sefiu Taiwo, Emmanuel Alayande College Of Education, Nigeria  
Mr. Sumit Goyal, National Dairy Research Institute, India  
Mr Jaswinder Singh Dilawari, Geeta Engineering College, India  
Prof. Raghuraj Singh, Harcourt Butler Technological Institute, Kanpur  
Dr. S.K. Mahendran, Anna University, Chennai, India  
Dr. Amit Wason, Hindustan Institute of Technology & Management, Punjab  
Dr. Ashu Gupta, Apeejay Institute of Management, India  
Assist. Prof. D. Asir Antony Gnana Singh, M.I.E.T Engineering College, India  
Mrs Mina Farmanbar, Eastern Mediterranean University, Famagusta, North Cyprus  
Mr. Maram Balajee, GMR Institute of Technology, India  
Mr. Moiz S. Ansari, Isra University, Hyderabad, Pakistan  
Mr. Adebayo, Olawale Surajudeen, Federal University of Technology Minna, Nigeria  
Mr. Jasvir Singh, University College Of Engg., India  
Mr. Vivek Tiwari, MANIT, Bhopal, India  
Assoc. Prof. R. Navaneethakrishnan, Bharathiyar College of Engineering and Technology, India  
Mr. Somdip Dey, St. Xavier's College, Kolkata, India  
Mr. Souleymane Balla-Arabé, Xi'an University of Electronic Science and Technology, China  
Mr. Mahabub Alam, Rajshahi University of Engineering and Technology, Bangladesh  
Mr. Sathyapraksh P., S.K.P Engineering College, India  
Dr. N. Karthikeyan, SNS College of Engineering, Anna University, India  
Dr. Binod Kumar, JSPM's, Jayawant Technical Campus, Pune, India  
Assoc. Prof. Dinesh Goyal, Suresh Gyan Vihar University, India  
Mr. Md. Abdul Ahad, K L University, India  
Mr. Vikas Bajpai, The LNM IIT, India  
Dr. Manish Kumar Anand, Salesforce (R & D Analytics), San Francisco, USA  
Assist. Prof. Dheeraj Murari, Kumaon Engineering College, India  
Assoc. Prof. Dr. A. Muthukumaravel, VELS University, Chennai  
Mr. A. Siles Balasingh, St. Joseph University in Tanzania, Tanzania  
Mr. Ravindra Daga Badgujar, R C Patel Institute of Technology, India  
Dr. Preeti Khanna, SVKM's NMIMS, School of Business Management, India  
Mr. Kumar Dayanand, Cambridge Institute of Technology, India  
Dr. Syed Asif Ali, SMI University Karachi, Pakistan  
Prof. Pallvi Pandit, Himachal Pradesh University, India  
Mr. Ricardo Verschueren, University of Gloucestershire, UK  
Assist. Prof. Mamta Juneja, University Institute of Engineering and Technology, Panjab University, India  
Assoc. Prof. P. Surendra Varma, NRI Institute of Technology, JNTU Kakinada, India  
Assist. Prof. Gaurav Shrivastava, RGPV / SVITS Indore, India  
Dr. S. Sumathi, Anna University, India

Assist. Prof. Ankita M. Kapadia, Charotar University of Science and Technology, India  
Mr. Deepak Kumar, Indian Institute of Technology (BHU), India  
Dr. Dr. Rajan Gupta, GGSIP University, New Delhi, India  
Assist. Prof M. Anand Kumar, Karpagam University, Coimbatore, India  
Mr. Mr Arshad Mansoor, Pakistan Aeronautical Complex  
Mr. Kapil Kumar Gupta, Ansal Institute of Technology and Management, India  
Dr. Neeraj Tomer, SINE International Institute of Technology, Jaipur, India  
Assist. Prof. Trunal J. Patel, C.G.Patel Institute of Technology, Uka Tarsadia University, Bardoli, Surat  
Mr. Sivakumar, Codework solutions, India  
Mr. Mohammad Sadegh Mirzaei, PGNR Company, Iran  
Dr. Gerard G. Dumancas, Oklahoma Medical Research Foundation, USA  
Mr. Varadala Sridhar, Varadhman College Engineering College, Affiliated To JNTU, Hyderabad  
Assist. Prof. Manoj Dhawan, SVITS, Indore  
Assoc. Prof. Chitreshh Banerjee, Suresh Gyan Vihar University, Jaipur, India  
Dr. S. Santhi, SCSVMV University, India  
Mr. Davood Mohammadi Souran, Ministry of Energy of Iran, Iran  
Mr. Shamim Ahmed, Bangladesh University of Business and Technology, Bangladesh  
Mr. Sandeep Reddivari, Mississippi State University, USA  
Assoc. Prof. Ousmane Thiare, Gaston Berger University, Senegal  
Dr. Hazra Imran, Athabasca University, Canada  
Dr. Setu Kumar Chaturvedi, Technocrats Institute of Technology, Bhopal, India  
Mr. Mohd Dilshad Ansari, Jaypee University of Information Technology, India  
Ms. Jaspreet Kaur, Distance Education LPU, India  
Dr. D. Nagarajan, Salalah College of Technology, Sultanate of Oman  
Dr. K.V.N.R.Sai Krishna, S.V.R.M. College, India  
Mr. Himanshu Pareek, Center for Development of Advanced Computing (CDAC), India  
Mr. Khaldi Amine, Badji Mokhtar University, Algeria  
Mr. Mohammad Sadegh Mirzaei, Scientific Applied University, Iran  
Assist. Prof. Khyati Chaudhary, Ram-eesh Institute of Engg. & Technology, India  
Mr. Sanjay Agal, Pacific College of Engineering Udaipur, India  
Mr. Abdul Mateen Ansari, King Khalid University, Saudi Arabia  
Dr. H.S. Behera, Veer Surendra Sai University of Technology (VSSUT), India  
Dr. Shrikant Tiwari, Shri Shankaracharya Group of Institutions (SSGI), India  
Prof. Ganesh B. Regulwar, Shri Shankarprasad Agnihotri College of Engg, India  
Prof. Pinnamaneni Bhanu Prasad, Matrix vision GmbH, Germany  
Dr. Shrikant Tiwari, Shri Shankaracharya Technical Campus (SSTC), India  
Dr. Siddesh G.K., : Dayananada Sagar College of Engineering, Bangalore, India  
Dr. Nadir Bouchama, CERIST Research Center, Algeria  
Dr. R. Sathishkumar, Sri Venkateswara College of Engineering, India  
Assistant Prof (Dr.) Mohamed Moussaoui, Abdelmalek Essaadi University, Morocco  
Dr. S. Malathi, Panimalar Engineering College, Chennai, India  
Dr. V. Subedha, Panimalar Institute of Technology, Chennai, India

Dr. Prashant Panse, Swami Vivekanand College of Engineering, Indore, India  
Dr. Hamza Aldabbas, Al-Balqa'a Applied University, Jordan  
Dr. G. Rasitha Banu, Vel's University, Chennai  
Dr. V. D. Ambeth Kumar, Panimalar Engineering College, Chennai  
Prof. Anuranjan Misra, Bhagwant Institute of Technology, Ghaziabad, India  
Ms. U. Sinthuja, PSG college of arts & science, India  
Dr. Ehsan Saradar Torshizi, Urmia University, Iran  
Dr. Shamneesh Sharma, APG Shimla University, Shimla (H.P.), India  
Assistant Prof. A. S. Syed Navaz, Muthayammal College of Arts & Science, India  
Assistant Prof. Ranjit Panigrahi, Sikkim Manipal Institute of Technology, Majitar, Sikkim  
Dr. Khaled Eskaf, Arab Academy for Science ,Technology & Maritime Transportation, Egypt  
Dr. Nishant Gupta, University of Jammu, India  
Assistant Prof. Nagarajan Sankaran, Annamalai University, Chidambaram, Tamilnadu, India  
Assistant Prof. Tribikram Pradhan, Manipal Institute of Technology, India  
Dr. Nasser Lotfi, Eastern Mediterranean University, Northern Cyprus  
Dr. R. Manavalan, K S Rangasamy college of Arts and Science, Tamilnadu, India  
Assistant Prof. P. Krishna Sankar, K S Rangasamy college of Arts and Science, Tamilnadu, India  
Dr. Rahul Malik, Cisco Systems, USA  
Dr. S. C. Lingareddy, ALPHA College of Engineering, India  
Assistant Prof. Mohammed Shuaib, Interl University, Lucknow, India  
Dr. Sachin Yele, Sanghvi Institute of Management & Science, India  
Dr. T. Thambidurai, Sun Univercell, Singapore  
Prof. Anandkumar Telang, BKIT, India  
Assistant Prof. R. Poorvadevi, SCSVMV University, India  
Dr Uttam Mande, Gitam University, India  
Dr. Poornima Girish Naik, Shahu Institute of Business Education and Research (SIBER), India  
Prof. Md. Abu Kausar, Jaipur National University, Jaipur, India  
Dr. Mohammed Zuber, AISECT University, India  
Prof. Kalum Priyanath Udagepola, King Abdulaziz University, Saudi Arabia  
Dr. K. R. Ananth, Velalar College of Engineering and Technology, India  
Assistant Prof. Sanjay Sharma, Roorkee Engineering & Management Institute Shamli (U.P), India  
Assistant Prof. Panem Charan Arur, Priyadarshini Institute of Technology, India  
Dr. Ashwak Mahmood muhsen alabaichi, Karbala University / College of Science, Iraq  
Dr. Urmila Shrawankar, G H Raisonni College of Engineering, Nagpur (MS), India  
Dr. Krishan Kumar Paliwal, Panipat Institute of Engineering & Technology, India  
Dr. Mukesh Negi, Tech Mahindra, India  
Dr. Anuj Kumar Singh, Amity University Gurgaon, India  
Dr. Babar Shah, Gyeongsang National University, South Korea  
Assistant Prof. Jayprakash Upadhyay, SRI-TECH Jabalpur, India  
Assistant Prof. Varadala Sridhar, Vidya Jyothi Institute of Technology, India  
Assistant Prof. Parameshachari B D, KSIT, Bangalore, India  
Assistant Prof. Ankit Garg, Amity University, Haryana, India

Assistant Prof. Rajashe Karappa, SDM CET, Karnataka, India  
Assistant Prof. Varun Jasuja, GNIT, India  
Assistant Prof. Sonal Honale, Abha Gaikwad Patil College of Engineering Nagpur, India  
Dr. Pooja Choudhary, CT Group of Institutions, NIT Jalandhar, India  
Dr. Faouzi Hidoussi, UHL Batna, Algeria  
Dr. Naseer Ali Husieen, Wasit University, Iraq  
Assistant Prof. Vinod Kumar Shukla, Amity University, Dubai  
Dr. Ahmed Farouk Metwaly, K L University  
Mr. Mohammed Noaman Murad, Cihan University, Iraq  
Dr. Suxing Liu, Arkansas State University, USA  
Dr. M. Gomathi, Velalar College of Engineering and Technology, India  
Assistant Prof. Sumardiono, College PGRI Blitar, Indonesia  
Dr. Latika Kharb, Jagan Institute of Management Studies (JIMS), Delhi, India  
Associate Prof. S. Raja, Pauls College of Engineering and Technology, Tamilnadu, India  
Assistant Prof. Seyed Reza Pakize, Shahid Sani High School, Iran  
Dr. Thiyagu Nagaraj, University-INOUE, India  
Assistant Prof. Noreen Sarai, Harare Institute of Technology, Zimbabwe  
Assistant Prof. Gajanand Sharma, Suresh Gyan Vihar University Jaipur, Rajasthan, India  
Assistant Prof. Mapari Vikas Prakash, Siddhant COE, Sudumbare, Pune, India  
Dr. Devesh Katiyar, Shri Ramswaroop Memorial University, India  
Dr. Shenshen Liang, University of California, Santa Cruz, US  
Assistant Prof. Mohammad Abu Omar, Limkokwing University of Creative Technology- Malaysia  
Mr. Snehasis Banerjee, Tata Consultancy Services, India  
Assistant Prof. Kibona Lusekelo, Ruaha Catholic University (RUCU), Tanzania  
Assistant Prof. Adib Kabir Chowdhury, University College Technology Sarawak, Malaysia  
Dr. Ying Yang, Computer Science Department, Yale University, USA  
Dr. Vinay Shukla, Institute Of Technology & Management, India  
Dr. Liviu Octavian Mafteiu-Scail, West University of Timisoara, Romania  
Assistant Prof. Rana Khudhair Abbas Ahmed, Al-Rafidain University College, Iraq  
Assistant Prof. Nitin A. Naik, S.R.T.M. University, India  
Dr. Timothy Powers, University of Hertfordshire, UK  
Dr. S. Prasath, Bharathiar University, Erode, India  
Dr. Ritu Shrivastava, SIRTHS Bhopal, India  
Prof. Rohit Shrivastava, Mittal Institute of Technology, Bhopal, India  
Dr. Gianina Mihai, Dunarea de Jos" University of Galati, Romania  
Assistant Prof. Ms. T. Kalai Selvi, Erode Sengunthar Engineering College, India  
Assistant Prof. Ms. C. Kavitha, Erode Sengunthar Engineering College, India  
Assistant Prof. K. Sinivasamoorthi, Erode Sengunthar Engineering College, India  
Assistant Prof. Mallikarjun C Sarsamba Bheemna Khandre Institute Technology, Bhalki, India  
Assistant Prof. Vishwanath Chikaraddi, Veermata Jijabai technological Institute (Central Technological Institute), India  
Assistant Prof. Dr. Ikvinderpal Singh, Trai Shatabdi GGS Khalsa College, India

Assistant Prof. Mohammed Noaman Murad, Cihan University, Iraq  
Professor Yousef Farhaoui, Moulay Ismail University, Errachidia, Morocco  
Dr. Parul Verma, Amity University, India  
Professor Yousef Farhaoui, Moulay Ismail University, Errachidia, Morocco  
Assistant Prof. Madhavi Dhingra, Amity University, Madhya Pradesh, India  
Assistant Prof.. G. Selvavinayagam, SNS College of Technology, Coimbatore, India  
Assistant Prof. Madhavi Dhingra, Amity University, MP, India  
Professor Kartheesan Log, Anna University, Chennai  
Professor Vasudeva Acharya, Shri Madhwa vadiraja Institute of Technology, India  
Dr. Asif Iqbal Hajamydeen, Management & Science University, Malaysia  
Assistant Prof., Mahendra Singh Meena, Amity University Haryana  
Assistant Professor Manjeet Kaur, Amity University Haryana  
Dr. Mohamed Abd El-Basset Matwalli, Zagazig University, Egypt  
Dr. Ramani Kannan, Universiti Teknologi PETRONAS, Malaysia  
Assistant Prof. S. Jagadeesan Subramaniam, Anna University, India  
Assistant Prof. Dharmendra Choudhary, Tripura University, India  
Assistant Prof. Deepika Vodnala, SR Engineering College, India  
Dr. Kai Cong, Intel Corporation & Computer Science Department, Portland State University, USA  
Dr. Kailas R Patil, Vishwakarma Institute of Information Technology (VIIT), India  
Dr. Omar A. Alzubi, Faculty of IT / Al-Balqa Applied University, Jordan  
Assistant Prof. Kareemullah Shaik, Nimra Institute of Science and Technology, India  
Assistant Prof. Chirag Modi, NIT Goa  
Dr. R. Ramkumar, Nandha Arts And Science College, India  
Dr. Priyadharshini Vydhialingam, Harathiar University, India  
Dr. P. S. Jagadeesh Kumar, DBIT, Bangalore, Karnataka  
Dr. Vikas Thada, AMITY University, Pachgaon  
Dr. T. A. Ashok Kumar, Institute of Management, Christ University, Bangalore  
Dr. Shaheera Rashwan, Informatics Research Institute  
Dr. S. Preetha Gunasekar, Bharathiyar University, India  
Asst Professor Sameer Dev Sharma, Uttaranchal University, Dehradun  
Dr. Zhihan Iv, Chinese Academy of Science, China  
Dr. Ikvinderpal Singh, Trai Shatabdi GGS Khalsa College, Amritsar  
Dr. Umar Ruhi, University of Ottawa, Canada  
Dr. Jasmin Cosic, University of Bihac, Bosnia and Herzegovina  
Dr. Homam Reda El-Taj, University of Tabuk, Kingdom of Saudi Arabia  
Dr. Mostafa Ghobaei Arani, Islamic Azad University, Iran  
Dr. Ayyasamy Ayyanar, Annamalai University, India  
Dr. Selvakumar Manickam, Universiti Sains Malaysia, Malaysia  
Dr. Murali Krishna Namana, GITAM University, India  
Dr. Smriti Agrawal, Chaitanya Bharathi Institute of Technology, Hyderabad, India  
Professor Vimalathithan Rathinasabapathy, Karpagam College Of Engineering, India  
Dr. Sushil Chandra Dimri, Graphic Era University, India



Dr. Dinh-Sinh Mai, Le Quy Don Technical University, Vietnam

Dr. S. Rama Sree, Aditya Engg. College, India

Dr. Ehab T. Alnfrawy, Sadat Academy, Egypt

Dr. Patrick D. Cerna, Haramaya University, Ethiopia

Dr. Vishal Jain, Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), India

# CALL FOR PAPERS

## International Journal of Computer Science and Information Security

IJCSIS 2015

ISSN: 1947-5500

<http://sites.google.com/site/ijcsis/>

International Journal Computer Science and Information Security, IJCSIS, is the premier scholarly venue in the areas of computer science and security issues. IJCSIS 2011 will provide a high profile, leading edge platform for researchers and engineers alike to publish state-of-the-art research in the respective fields of information technology and communication security. The journal will feature a diverse mixture of publication articles including core and applied computer science related topics.

Authors are solicited to contribute to the special issue by submitting articles that illustrate research results, projects, surveying works and industrial experiences that describe significant advances in the following areas, but are not limited to. Submissions may span a broad range of topics, e.g.:

### *Track A: Security*

Access control, Anonymity, Audit and audit reduction & Authentication and authorization, Applied cryptography, Cryptanalysis, Digital Signatures, Biometric security, Boundary control devices, Certification and accreditation, Cross-layer design for security, Security & Network Management, Data and system integrity, Database security, Defensive information warfare, Denial of service protection, Intrusion Detection, Anti-malware, Distributed systems security, Electronic commerce, E-mail security, Spam, Phishing, E-mail fraud, Virus, worms, Trojan Protection, Grid security, Information hiding and watermarking & Information survivability, Insider threat protection, Integrity  
Intellectual property protection, Internet/Intranet Security, Key management and key recovery, Language-based security, Mobile and wireless security, Mobile, Ad Hoc and Sensor Network Security, Monitoring and surveillance, Multimedia security ,Operating system security, Peer-to-peer security, Performance Evaluations of Protocols & Security Application, Privacy and data protection, Product evaluation criteria and compliance, Risk evaluation and security certification, Risk/vulnerability assessment, Security & Network Management, Security Models & protocols, Security threats & countermeasures (DDoS, MiM, Session Hijacking, Replay attack etc.), Trusted computing, Ubiquitous Computing Security, Virtualization security, VoIP security, Web 2.0 security, Submission Procedures, Active Defense Systems, Adaptive Defense Systems, Benchmark, Analysis and Evaluation of Security Systems, Distributed Access Control and Trust Management, Distributed Attack Systems and Mechanisms, Distributed Intrusion Detection/Prevention Systems, Denial-of-Service Attacks and Countermeasures, High Performance Security Systems, Identity Management and Authentication, Implementation, Deployment and Management of Security Systems, Intelligent Defense Systems, Internet and Network Forensics, Large-scale Attacks and Defense, RFID Security and Privacy, Security Architectures in Distributed Network Systems, Security for Critical Infrastructures, Security for P2P systems and Grid Systems, Security in E-Commerce, Security and Privacy in Wireless Networks, Secure Mobile Agents and Mobile Code, Security Protocols, Security Simulation and Tools, Security Theory and Tools, Standards and Assurance Methods, Trusted Computing, Viruses, Worms, and Other Malicious Code, World Wide Web Security, Novel and emerging secure architecture, Study of attack strategies, attack modeling, Case studies and analysis of actual attacks, Continuity of Operations during an attack, Key management, Trust management, Intrusion detection techniques, Intrusion response, alarm management, and correlation analysis, Study of tradeoffs between security and system performance, Intrusion tolerance systems, Secure protocols, Security in wireless networks (e.g. mesh networks, sensor networks, etc.), Cryptography and Secure Communications, Computer Forensics, Recovery and Healing, Security Visualization, Formal Methods in Security, Principles for Designing a Secure Computing System, Autonomic Security, Internet Security, Security in Health Care Systems, Security Solutions Using Reconfigurable Computing, Adaptive and Intelligent Defense Systems, Authentication and Access control, Denial of service attacks and countermeasures, Identity, Route and

Location Anonymity schemes, Intrusion detection and prevention techniques, Cryptography, encryption algorithms and Key management schemes, Secure routing schemes, Secure neighbor discovery and localization, Trust establishment and maintenance, Confidentiality and data integrity, Security architectures, deployments and solutions, Emerging threats to cloud-based services, Security model for new services, Cloud-aware web service security, Information hiding in Cloud Computing, Securing distributed data storage in cloud, Security, privacy and trust in mobile computing systems and applications, **Middleware security & Security features:** middleware software is an asset on its own and has to be protected, interaction between security-specific and other middleware features, e.g., context-awareness, **Middleware-level security monitoring and measurement:** metrics and mechanisms for quantification and evaluation of security enforced by the middleware, **Security co-design:** trade-off and co-design between application-based and middleware-based security, **Policy-based management:** innovative support for policy-based definition and enforcement of security concerns, **Identification and authentication mechanisms:** Means to capture application specific constraints in defining and enforcing access control rules, **Middleware-oriented security patterns:** identification of patterns for sound, reusable security, **Security in aspect-based middleware:** mechanisms for isolating and enforcing security aspects, **Security in agent-based platforms:** protection for mobile code and platforms, Smart Devices: Biometrics, National ID cards, Embedded Systems Security and TPMs, RFID Systems Security, Smart Card Security, Pervasive Systems: Digital Rights Management (DRM) in pervasive environments, Intrusion Detection and Information Filtering, Localization Systems Security (Tracking of People and Goods), Mobile Commerce Security, Privacy Enhancing Technologies, Security Protocols (for Identification and Authentication, Confidentiality and Privacy, and Integrity), Ubiquitous Networks: Ad Hoc Networks Security, Delay-Tolerant Network Security, Domestic Network Security, Peer-to-Peer Networks Security, Security Issues in Mobile and Ubiquitous Networks, Security of GSM/GPRS/UMTS Systems, Sensor Networks Security, Vehicular Network Security, Wireless Communication Security: Bluetooth, NFC, WiFi, WiMAX, WiMedia, others

This Track will emphasize the design, implementation, management and applications of computer communications, networks and services. Topics of mostly theoretical nature are also welcome, provided there is clear practical potential in applying the results of such work.

### ***Track B: Computer Science***

Broadband wireless technologies: LTE, WiMAX, WiRAN, HSDPA, HSUPA, Resource allocation and interference management, Quality of service and scheduling methods, Capacity planning and dimensioning, Cross-layer design and Physical layer based issue, Interworking architecture and interoperability, Relay assisted and cooperative communications, Location and provisioning and mobility management, Call admission and flow/congestion control, Performance optimization, Channel capacity modeling and analysis, Middleware Issues: Event-based, publish/subscribe, and message-oriented middleware, Reconfigurable, adaptable, and reflective middleware approaches, Middleware solutions for reliability, fault tolerance, and quality-of-service, Scalability of middleware, Context-aware middleware, Autonomic and self-managing middleware, Evaluation techniques for middleware solutions, Formal methods and tools for designing, verifying, and evaluating, middleware, Software engineering techniques for middleware, Service oriented middleware, Agent-based middleware, Security middleware, Network Applications: Network-based automation, Cloud applications, Ubiquitous and pervasive applications, Collaborative applications, RFID and sensor network applications, Mobile applications, Smart home applications, Infrastructure monitoring and control applications, Remote health monitoring, GPS and location-based applications, Networked vehicles applications, Alert applications, Embedded Computer System, Advanced Control Systems, and Intelligent Control : Advanced control and measurement, computer and microprocessor-based control, signal processing, estimation and identification techniques, application specific IC's, nonlinear and adaptive control, optimal and robot control, intelligent control, evolutionary computing, and intelligent systems, instrumentation subject to critical conditions, automotive, marine and aero-space control and all other control applications, Intelligent Control System, Wiring/Wireless Sensor, Signal Control System. Sensors, Actuators and Systems Integration : Intelligent sensors and actuators, multisensor fusion, sensor array and multi-channel processing, micro/nano technology, microsensors and microactuators, instrumentation electronics, MEMS and system integration, wireless sensor, Network Sensor, Hybrid

Sensor, Distributed Sensor Networks. Signal and Image Processing : Digital signal processing theory, methods, DSP implementation, speech processing, image and multidimensional signal processing, Image analysis and processing, Image and Multimedia applications, Real-time multimedia signal processing, Computer vision, Emerging signal processing areas, Remote Sensing, Signal processing in education. Industrial Informatics: Industrial applications of neural networks, fuzzy algorithms, Neuro-Fuzzy application, bioInformatics, real-time computer control, real-time information systems, human-machine interfaces, CAD/CAM/CAT/CIM, virtual reality, industrial communications, flexible manufacturing systems, industrial automated process, Data Storage Management, Harddisk control, Supply Chain Management, Logistics applications, Power plant automation, Drives automation. Information Technology, Management of Information System : Management information systems, Information Management, Nursing information management, Information System, Information Technology and their application, Data retrieval, Data Base Management, Decision analysis methods, Information processing, Operations research, E-Business, E-Commerce, E-Government, Computer Business, Security and risk management, Medical imaging, Biotechnology, Bio-Medicine, Computer-based information systems in health care, Changing Access to Patient Information, Healthcare Management Information Technology. Communication/Computer Network, Transportation Application : On-board diagnostics, Active safety systems, Communication systems, Wireless technology, Communication application, Navigation and Guidance, Vision-based applications, Speech interface, Sensor fusion, Networking theory and technologies, Transportation information, Autonomous vehicle, Vehicle application of affective computing, Advance Computing technology and their application : Broadband and intelligent networks, Data Mining, Data fusion, Computational intelligence, Information and data security, Information indexing and retrieval, Information processing, Information systems and applications, Internet applications and performances, Knowledge based systems, Knowledge management, Software Engineering, Decision making, Mobile networks and services, Network management and services, Neural Network, Fuzzy logics, Neuro-Fuzzy, Expert approaches, Innovation Technology and Management : Innovation and product development, Emerging advances in business and its applications, Creativity in Internet management and retailing, B2B and B2C management, Electronic transceiver device for Retail Marketing Industries, Facilities planning and management, Innovative pervasive computing applications, Programming paradigms for pervasive systems, Software evolution and maintenance in pervasive systems, Middleware services and agent technologies, Adaptive, autonomic and context-aware computing, Mobile/Wireless computing systems and services in pervasive computing, Energy-efficient and green pervasive computing, Communication architectures for pervasive computing, Ad hoc networks for pervasive communications, Pervasive opportunistic communications and applications, Enabling technologies for pervasive systems (e.g., wireless BAN, PAN), Positioning and tracking technologies, Sensors and RFID in pervasive systems, Multimodal sensing and context for pervasive applications, Pervasive sensing, perception and semantic interpretation, Smart devices and intelligent environments, Trust, security and privacy issues in pervasive systems, User interfaces and interaction models, Virtual immersive communications, Wearable computers, Standards and interfaces for pervasive computing environments, Social and economic models for pervasive systems, Active and Programmable Networks, Ad Hoc & Sensor Network, Congestion and/or Flow Control, Content Distribution, Grid Networking, High-speed Network Architectures, Internet Services and Applications, Optical Networks, Mobile and Wireless Networks, Network Modeling and Simulation, Multicast, Multimedia Communications, Network Control and Management, Network Protocols, Network Performance, Network Measurement, Peer to Peer and Overlay Networks, Quality of Service and Quality of Experience, Ubiquitous Networks, Crosscutting Themes – Internet Technologies, Infrastructure, Services and Applications; Open Source Tools, Open Models and Architectures; Security, Privacy and Trust; Navigation Systems, Location Based Services; Social Networks and Online Communities; ICT Convergence, Digital Economy and Digital Divide, Neural Networks, Pattern Recognition, Computer Vision, Advanced Computing Architectures and New Programming Models, Visualization and Virtual Reality as Applied to Computational Science, Computer Architecture and Embedded Systems, Technology in Education, Theoretical Computer Science, Computing Ethics, Computing Practices & Applications

Authors are invited to submit papers through e-mail [ijcsiseditor@gmail.com](mailto:ijcsiseditor@gmail.com). Submissions must be original and should not have been published previously or be under consideration for publication while being evaluated by IJCSIS. Before submission authors should carefully read over the journal's Author Guidelines, which are located at <http://sites.google.com/site/ijcsis/authors-notes> .



**© IJCSIS PUBLICATION 2015**

**ISSN 1947 5500**

**<http://sites.google.com/site/ijcsis/>**